

# Leveraging orientation knowledge to enhance human pose estimation methods

**S. Azrour, S. Piérard, M. Van Droogenbroeck**

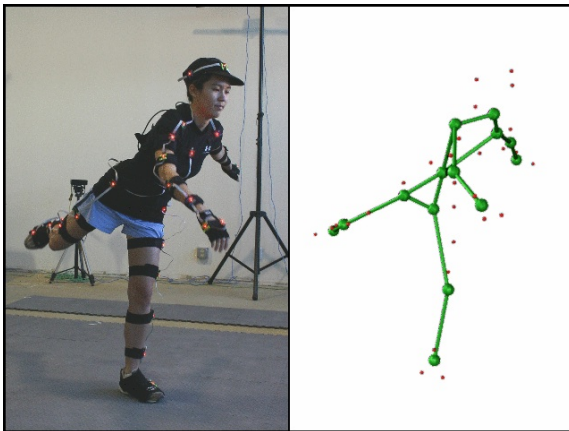
INTELSIG Laboratory, University of Liège, Belgium

Conference on Articulated Motion and Deformable Objects (AMDO 2016)  
13-15th July 2016

# What is human pose estimation ?

## Definition (Human pose estimation)

In computer vision, it is the study of algorithms and systems that recover the pose of a human body, which consists of joints and rigid parts.



# Application of human pose estimation: some examples

## Motion analysis



## Medical



## Entertainment



## Animation movies



# Types of camera-based pose estimation

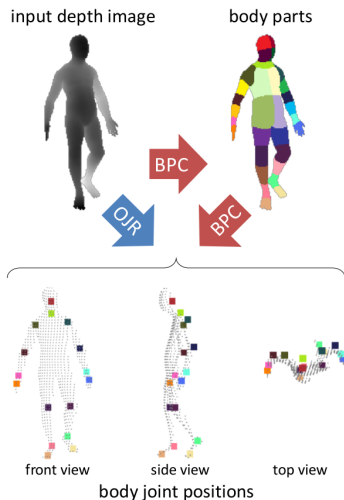
The camera-based pose estimation (or motion capture) can be **marker-based** or **markerless**:

**marker-based**: markers are put on the subject and the pose is recovered by localizing these markers with a multi-camera setup.

**markerless**: the subject has nothing to wear and its pose is recovered using a body model tracking method or a machine learning technique.

# Markerless pose estimation using a machine learning technique

- Pose estimation algorithms developed by Microsoft for the Kinect camera (from "J. Shotton, R. Girshick *et al.*, PAMI 2013").



# Silhouette ambiguity

- ▶ There is an intrinsic limitation when using color cameras: for one given silhouette, two different poses are possible



=



or



?

⇒ Depth cameras help to overcome this limitation but it still remains hard to disambiguate the silhouette orientation and predict the body joint positions at the same time.

# Using an orientation information to improve the pose estimation

## Idea

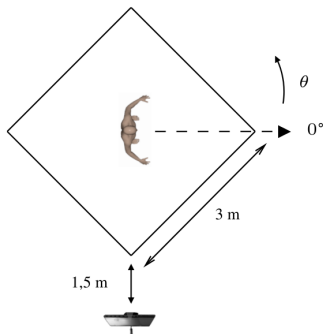
It is preferable to rely on an additional method that is specifically designed for orientation estimation instead of trying to recover the joint positions and disambiguate the silhouette orientation all at once.

How can we estimate the orientation ?

- ▶ The orientation estimation can be obtained from the image itself or thanks to any kind of sensors through a machine learning or a tracking algorithm.

# Using an orientation information to improve the pose estimation

- The configuration considered in this work:



- How do we take advantage of the orientation estimation ?  
 $\implies$  We slice the full orientation range into smaller ranges and learn a different model for each of these smaller ranges.

# Outline of our method

3D camera (e.g. Kinect)



or

Range Laser Scanner



or

...



Tracking  
or  
machine learning



Orientation estimation



Select right model



Set of Machine Learning  
models specialized for  
different ranges of  
orientations



...

1

2

3



Depth image



Right model



n

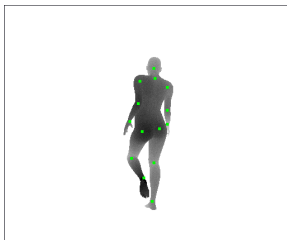
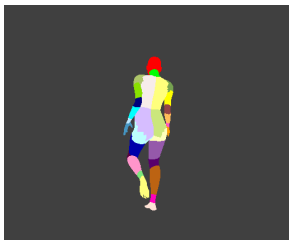
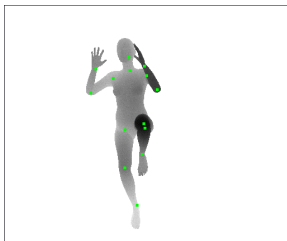
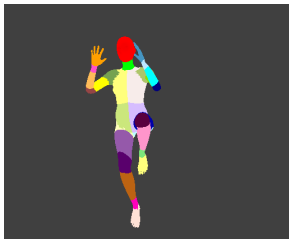


Pose prediction



# Synthetic data generation

- ▶ The body model is created with **MakeHuman**.
- ▶ Depth images are rendered inside **Blender**.
- ▶ Poses are taken randomly from the **CMU motion capture database**.



We use our own implementation of the *offset joint regression* algorithm proposed by Microsoft (R. Girshick et al., ICCV, 2011).

- ▶ The machine learning technique used is a random forest.
- ▶ Each pixel of the silhouette predicts a set of 3D offsets toward the body joints.
- ▶ These predictions are then aggregated using Mean Shift.

- ▶ We compared the accuracy of the estimated pose when using 1, 4 and 12 models.
- ▶ We considered two scenarios:
  - 1 A constant *global* learning dataset size.
  - 2 A constant learning dataset size *per model*.

# Results with a constant *global* learning dataset size

		amount of models:	1		4		12
		learning samples per model:	8000		$8000/4 = 2000$		$8000/12 \simeq 666$
		range of each model:	$360^\circ$		$360^\circ/4 + 2 \times 10^\circ = 110^\circ$		$360^\circ/12 + 2 \times 10^\circ = 50^\circ$
mean error	neck	2.9 cm	>		2.4 cm (- 15.3 %)	<	2.4 cm (- 14.4 %)
	head	3.1 cm	>		2.8 cm (- 7.6 %)	<	2.9 cm (- 3.9 %)
	right shoulder	5.4 cm	>		3.0 cm (- 45.1 %)	<	3.0 cm (- 44.1 %)
	right elbow	9.1 cm	>		5.9 cm (- 35.3 %)	<	6.0 cm (- 34.0 %)
	right wrist	13.7 cm	>		9.9 cm (- 27.3 %)	<	10.3 cm (- 24.4 %)
	right hip	4.2 cm	>		2.8 cm (- 34.0 %)	<	2.8 cm (- 33.6 %)
	right knee	5.8 cm	>		4.5 cm (- 23.4 %)	<	4.6 cm (- 21.8 %)
	right ankle	8.3 cm	>		6.2 cm (- 25.5 %)	<	6.3 cm (- 23.9 %)

⇒ Significant reduction of the error when going from 1 to 4 models.

⇒ However, going from 4 to 12 models slightly worsens the performance.

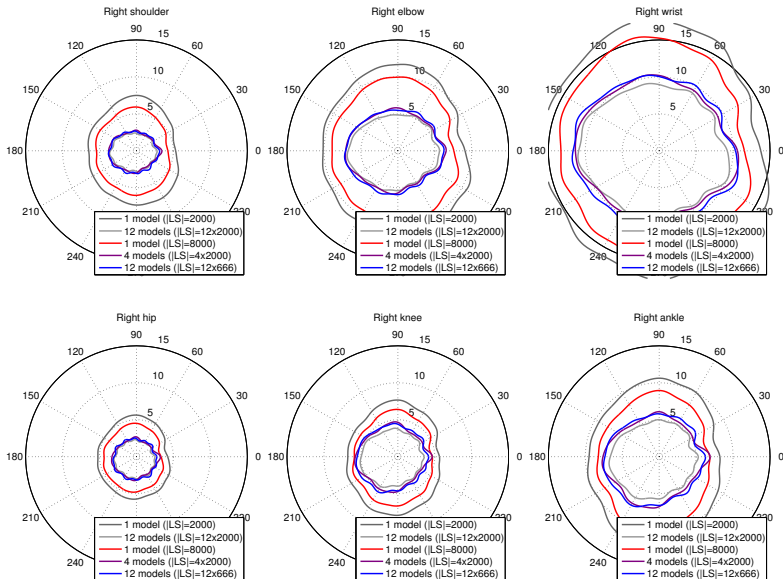
# Results with a constant learning dataset size *per model*

amount of models:		1		4		12
learning samples per model:		2000		2000		2000
range of each model:		360°		$360^\circ/4 + 2 \times 10^\circ = 110^\circ$		$360^\circ/12 + 2 \times 10^\circ = 50^\circ$
mean error	neck	3.3 cm	>	2.4 cm (- 27.2 %)	>	2.2 cm (- 34.4 %)
	head	3.5 cm	>	2.8 cm (- 19.8 %)	>	2.6 cm (- 25.3 %)
	right shoulder	6.6 cm	>	3.0 cm (- 55.4 %)	>	2.7 cm (- 58.7 %)
	right elbow	10.6 cm	>	5.9 cm (- 44.6 %)	>	5.4 cm (- 48.5 %)
	right wrist	15.9 cm	>	9.9 cm (- 37.6 %)	>	9.3 cm (- 41.7 %)
	right hip	5.2 cm	>	2.8 cm (- 46.0 %)	>	2.6 cm (- 50.1 %)
	right knee	6.9 cm	>	4.5 cm (- 35.6 %)	>	3.9 cm (- 43.0 %)
	right ankle	9.9 cm	>	6.2 cm (- 37.4 %)	>	5.4 cm (- 45.3 %)

⇒ Systematic decrease of the error when the number of models is increased.

⇒ However, small difference between 4 and 12 models suggests a plateau is reached.

# Mean error according to the orientation



- ▶ We can improve the accuracy of the estimated pose by taking advantage of an orientation estimation.
- ▶ One way to take advantage of the orientation estimation is to learn multiple models specialized for different range of orientations.
- ▶ We show that accuracy can be significantly improved when the number of models increases, even while keeping a constant global learning dataset size.