Improving pose estimation by building dedicated datasets and using orientation

Samir Azrour
Samir.Azrour@ulg.ac.be

Sébastien Piérard
Sebastien.Pierard@ulg.ac.be

Marc Van Droogenbroeck
M.VanDroogenbroeck@ulg.ac.be

INTELSIG, Montefiore Institute, University of Liège, Belgium

Background and objective

Markerless pose estimation systems are useful for various applications including human-computer interaction, activity recognition, security, gait analysis, and computer-assisted medical interventions. They have attracted much interest since the release of low-cost depth cameras such as Microsoft's Kinect camera. Shotton et al. and Girshick et al. pioneered tractable methods that infer a full-body pose reconstruction in real-time. Details of these methods are given in [1].

Despite this technological breakthrough, the accuracy of human pose estimation from single depth images remains insufficient for some applications. Our work aims at building a simulation environment to create images databases suited for any camera position and improving the mainstream machine learning-based pose estimation algorithms.

Database

We are working on a simulation environment using the open source softwares Blender [2] and MakeHuman [3]. This simulation environment is used to generate the learning and test images for human pose estimation algorithms.

To generate those images, we followed a method similar to the one described in [1]. With our simulation environment, we generate depth images for any camera position and we can use different human models with various morphologies.

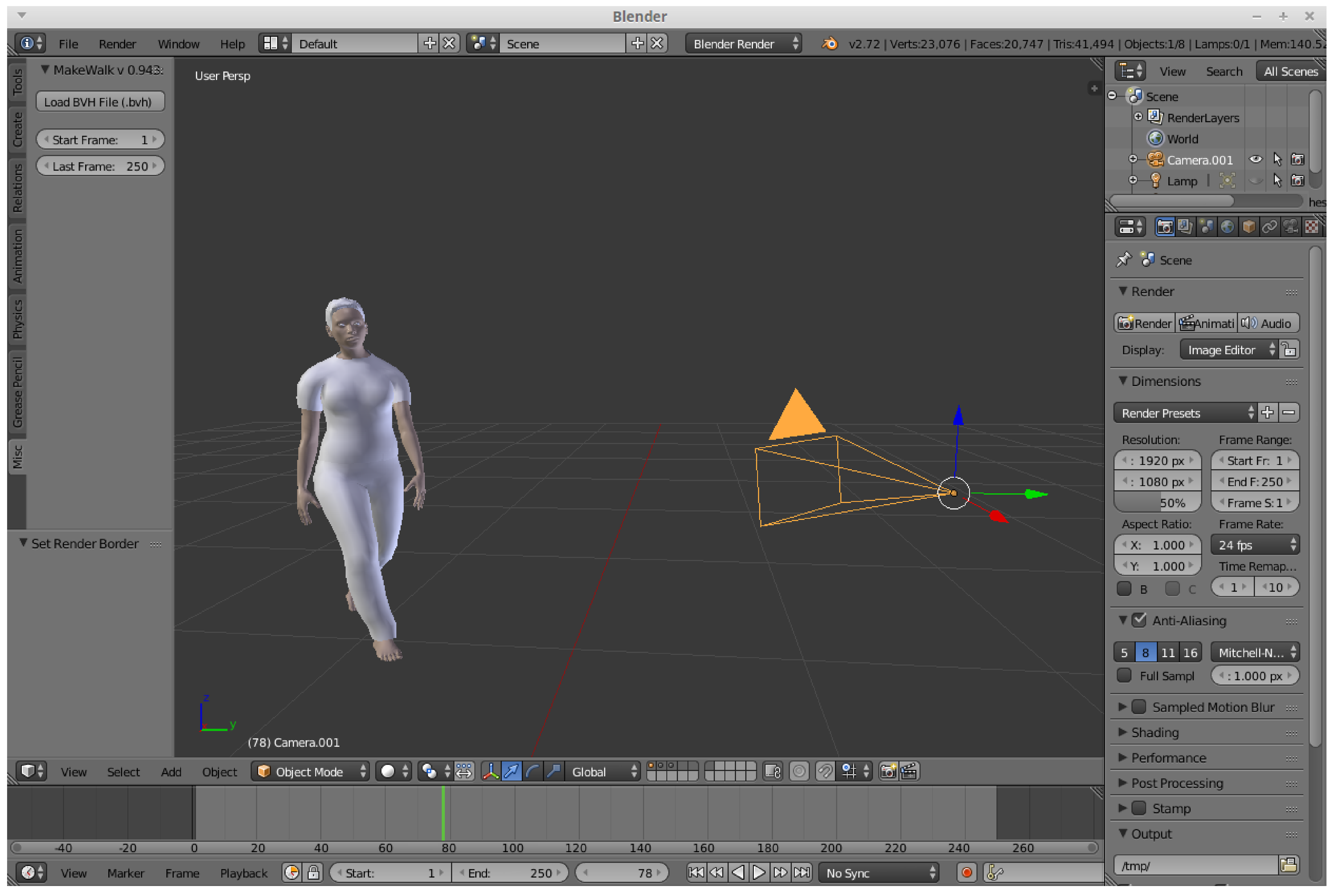

Figure 1 : The body model is created with MakeHuman [3] and can be animated using BVH (Biovision Hierarchy) files inside Blender [2]. The camera can be placed in any position and orientation. Color as well as depth images can be rendered.

There are several ways to build a set of poses. The poses can be obtained with a marker-based motion capture systems (as the Vicon system for instance [4]). Some databases of real poses, like the CMU motion capture database [5], are freely available. As an alternative, the poses can also be generated artificially as done in [6].

The main advantage of a simulation environment is that it can be tuned to the real camera position and appropriate human models to improve the accuracy. Moreover, it can be used to determine the best camera position for a given application.

The information associated with the generated images are the 3D body joints positions and the body part label for each pixel of the silhouette. Some generated images, with the associated informations, are shown in the figure below.

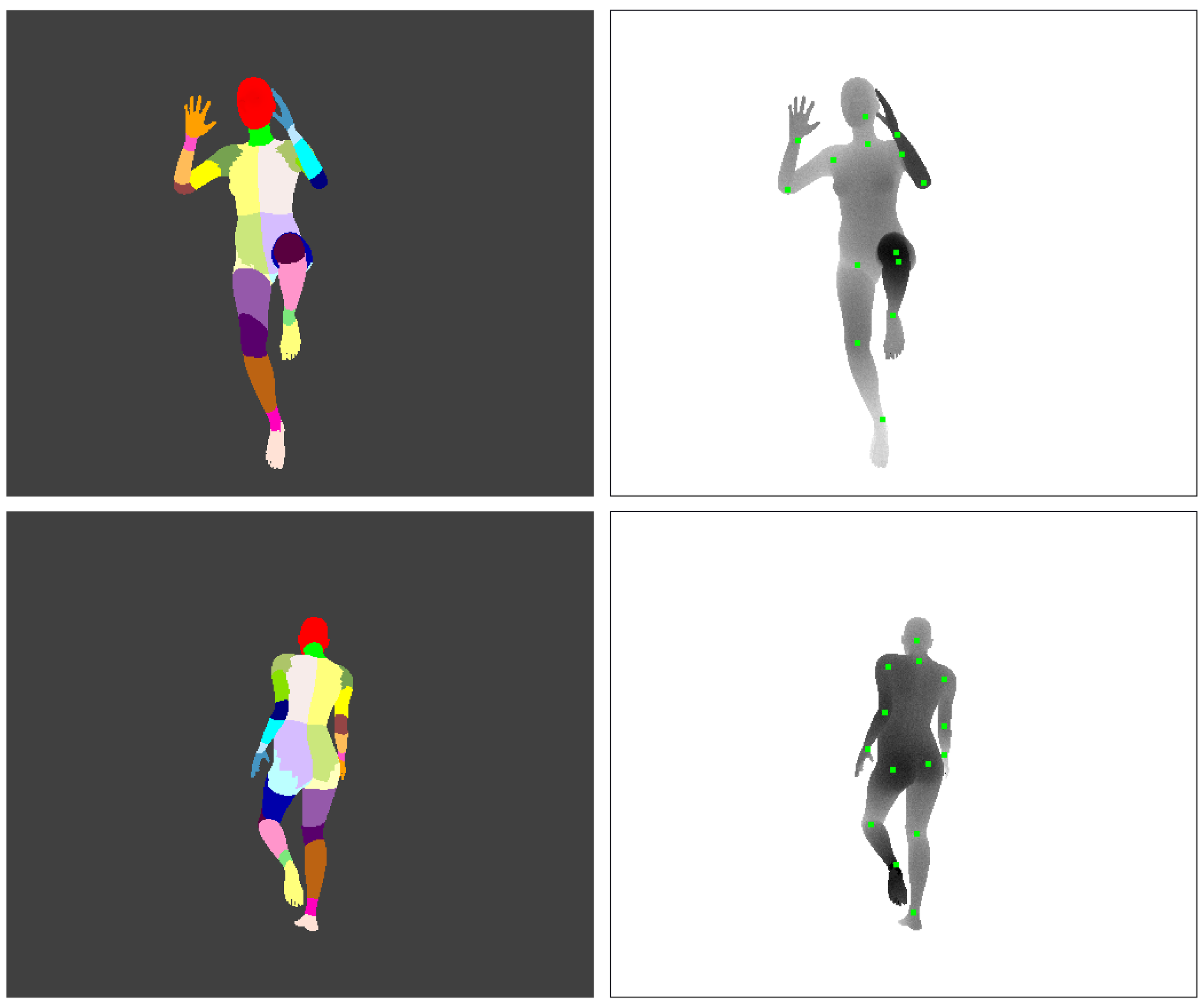

Figure 2 : Examples of generated depth images and corresponding body part label images. The ground truth body joints are displayed in green in the depth images.

First results by considering the body orientation

We are investigating several ways to improve the machine learning-based human pose estimation method proposed in [1]. The improvements concerns several parts of the method. Currently, we concentrate on providing useful information about the person being observed to the machine learning algorithm, in order to facilitate its task.

An example of such an information is the orientation of the observed person with respect to the camera which can be obtained from the image itself as in [7] or from any other clue like tracking. If we do not restrict the orientation of the observed person, the machine learning algorithm has to simultaneously estimate the orientation and the pose. By providing an orientation estimation, the task of the machine learning algorithm is eased because it can focus on the pose estimation. The first results show a significant improvement when an orientation estimation is provided to a pose estimation method. The figure below illustrates how we take into account the orientation information in our method.

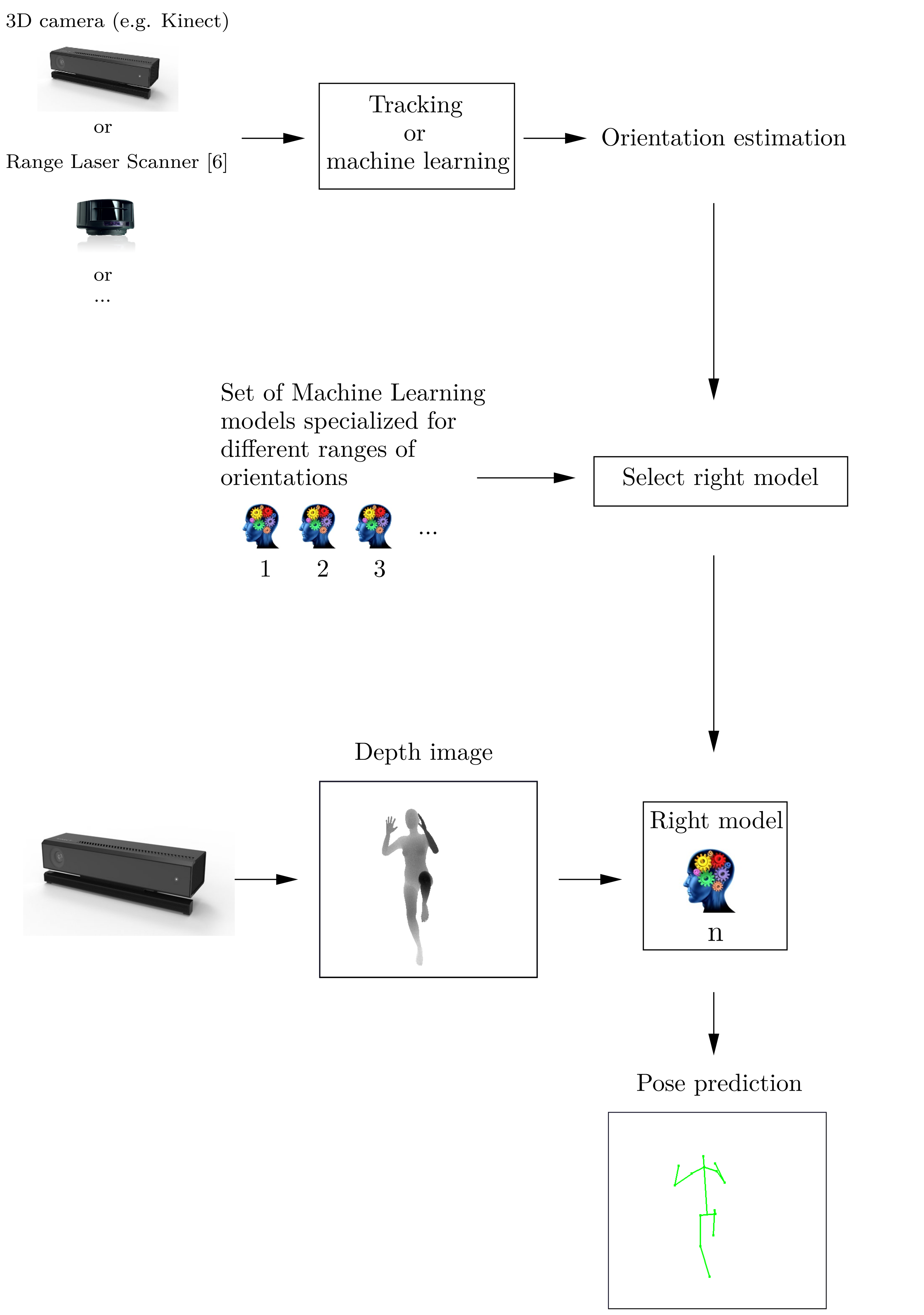

Figure 3 : Outline of our methodology. We use an orientation information about the observed person to improve the pose estimation. For each input depth image, we select the machine learning model learned over the appropriate range of orientations.

References

[1] J. Shotton, R. Girshick, A. Fitzgibbon, T. Sharp, M. Cook, M. Finocchio, R. Moore, P. Kohli, A. Criminisi, A. Kipman, and A. Blake. Efficient human pose estimation from single depth images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(12):2821–2840, December 2013.

[2] The Blender team. The Blender website. http://www.blender.org.

[3] The MakeHuman team. The MakeHuman website. http://www.makehuman.org.

[4] Vicon Motion Systems. http://www.vicon.com.

[5] Carnegie Mellon University. Motion capture database. http://mocap.cs.cmu.edu.

[6] S. Piérard and M. Van Droogenbroeck. A technique for building databases of annotated and realistic human silhouettes based on an avatar. In *Workshop on Circuits, Systems and Signal Processing (ProRISC)*, pages 243–246, Veldhoven, The Netherlands, November 2009.

[7] S. Piérard, D. Leroy, J.-F. Hansen, and M. Van Droogenbroeck. Estimation of human orientation in images captured with a range camera. In *Advanced Concepts for Intelligent Vision Systems (ACIVS)*, volume 6915 of *Lecture Notes in Computer Science*, pages 519–530. Springer, 2011.

Acknowledgements. Samir Azrour is supported by a research fellowship of the Belgian National Fund for Scientific Research (F.R.S.-FNRS). Sébastien Piérard works on a project funded by the Walloon region of Belgium.

Disclosure information. The authors have nothing to disclose