Pattern Recognition Letters 56 (2015) 14–21

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Multi-part body segmentation based on depth maps for soft biometry analysis ☆

Meysam Madadi [a,*], Sergio Escalera [a,b], Jordi Gonzàlez [a,c], F. Xavier Roca [a,c], Felipe Lumbreras [a,c]

[a] Computer Vision Center, Campus UAB, Edifici O, 08193, Bellaterra, Spain
[b] Dept. Matemàtica Aplicada i Anàlisi, Universitat de Barcelona, Gran Via de les Corts Catalanes 585, 08007, Barcelona, Spain
[c] Dept. Computer Science, Universitat Autònoma de Barcelona, Edifici Q, 08193, Bellaterra, Barcelona, Spain

ARTICLE INFO

Article history:
Received 11 March 2014
Available online 2 February 2015

Keywords:
3D shape context
3D point cloud alignment
Depth maps
Human body segmentation
Soft biometry analysis

ABSTRACT

This paper presents a novel method extracting biometric measures using depth sensors. Given a multi-part labeled training data, a new subject is aligned to the best model of the dataset, and soft biometrics such as lengths or circumference sizes of limbs and body are computed. The process is performed by training relevant pose clusters, defining a representative model, and fitting a 3D shape context descriptor within an iterative matching procedure. We show robust measures by applying orthogonal plates to body hull. We test our approach in a novel full-body RGB-Depth data set, showing accurate estimation of soft biometrics and better segmentation accuracy in comparison with random forest approach without requiring large training data.

1. Introduction

Soft biometrics in contrast to hard biometrics are traits of the human body, like color of the hair, skin, height and weight, that can be used to describe a person. These attributes have a lower power to discriminate and authenticate an individual, but they are easier to compute in comparison to hard biometrics.

Soft biometric traits have been used in video surveillance to track people with single camera systems or even with a discrete joint camera network [1–3]; as a preprocessing approach to help hard biometric systems to search databases faster or to increase reliability and accuracy [4,5]; and for other applications like person re-identification [6], supported diagnosis in clinical setups [7], or commercial tools like clothing sizing [8], just to mention a few. Most surveillance systems using soft biometrics have integrated human height as one of their most important cues [2,3,9].

Velardo et al. [10] proposed a weight estimation technique that computes weight by summation of coefficients of some soft biometrics like height and calf circumference. Since soft biometrics have semantic correlation in human metrology, these can be computed according to part relations. Recently, Adjeroh et al. [11] studied the problem of predictability and correlation in human metrology applying some statistical measurements between different soft biometrics features in order to make correlation clusters among them to predict unknown body measurements. Samejima et al. [12] used joints estimated by KinectSDK to estimate initial dimensions, afterwards multiple Regression of the 2 principal components of estimated body dimensions were applied to estimate other dimensions. Weiss et al. [13] computed body measurements using a regression based approach from body parameters after an accurate scanning of the body.

Indeed the extraction of human body part traits in soft biometric systems, as other areas in computer vision, suffers from difficulties like illumination changes, cluttered and uncontrolled environments, and the fact of dealing with the articulated nature of the human body. Recently, Microsoft-Corp. [14] has launched a low price multi-sensor device that uses pseudo random structured light technology that is capable of capturing RGB images and depth information simultaneously, which makes it possible to acquire 3D coordinates of pixels with high accuracy in indoor environments and overcome most of the difficulties aforementioned.

While most of the biometrics measurements are based on regression on some known body parameters, in this paper, first we accurately segment human limbs from a single depth image captured by a Kinect camera, and as a result we compute traits such as arm and leg lengths, and neck, chest, stomach, waist and hip sizes from segmented limbs. We use Kinect to get the human point clouds using background subtraction and depth thresholding from real user data, see in Fig. 1 a typical pose, depth image, and the corresponding segments. As a first stage, we focus on human pose estimation as a multi-limb segmentation problem [15]. Two general approaches are defined for this task: model based and model free techniques. In model based approaches, a kinematic model approximates the shape of the body from measurements that best fit the observed image features [16–19].

☆ This paper has been recommended for acceptance by L. Yin.

* Corresponding author. Tel.: +34 692 259 778.
E-mail address: meysam.madadi@gmail.com, mmadadi@cvc.uab.es (M. Madadi).

http://dx.doi.org/10.1016/j.patrec.2015.01.012
0167-8655/

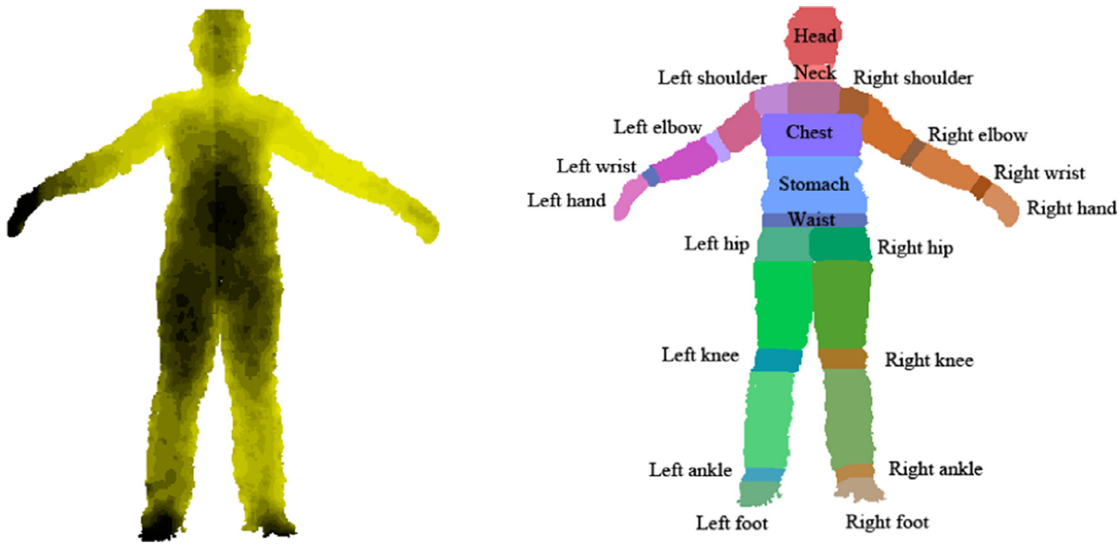

Fig. 1. A typical depth image and defined segments.

Andriluka et al. [20] successfully combined bottom-up part-based models with 3D top-down priors and showed the models capable to deal with more complex poses. Ramanan [21] proposed an edge based deformable model learned by a conditional random field, and used an iterative parsing as an energy minimization function to improve recognition. Several works are based on this approach as a first stage for human pose recovery and behavior analysis applications [22]. Recently, Ye et al. [23] used an example-based approach which finds the pose from the nearest sample after registration.

The methods for human pose estimation based on depth data have mainly focused on model free approaches. Model free approaches use feature vectors to learn and map feature space to pose space. Recently, Shotton et al. [24] proposed a random forest based approach to learn pixel labels from depth offsets, achieving robust segmentation results. This method has become one of the standard techniques for segmentation in depth data. However, this approach requires a huge dataset of real and synthetic labeled images as well as an expensive training procedure. Different works have focused on such a random forest segmentation approach to improve recognition of human body parts. Hernandez-Vela et al. [25] applied graph cuts to perform a local and spatial optimization of random forest output probabilities in order to improve segmentation accuracy. Kohli et al. [26] proposed a conditional regression forests approach applying a global latent variable that incorporates dependency between output variables, increasing body joint prediction.

In this work we use a model based system where labels of pixels are computed from a defined model after 3D alignment with the objective of performing soft biometrics analysis. For this task, we extract a depth image of each frame in the training set, and compute HOG features [27–29]. The described data is clustered to group similar poses in the same class in order to find the closest model to the test sample as fast as possible at test time instead of searching all the data set. The number of clusters is defined using a Gaussian mixture in an EM algorithm. With such an optimization, we are able to accurately cluster training data in a problem-dependent way without the need of prefixing clustering parameters.

Subsequently, the model is aligned to the test body sample in the 3D space using 3D shape context descriptors and 3D thin plate spline (TPS). Using HOG as a pose recognizer does not require 3D shape context to be invariant to rotation or viewpoint changes, although 3D shape context can be rotated based on eigenvectors of the point cloud. For our task we apply [30] 3D shape context for aligning point clouds of body hulls. In our procedure, a random number of pixels is selected and refined, removing nearest adjacent points, and then an iterative process finds the best matching points. Moreover each pixel in the body gets the nearest pixel label in the aligned model. As a result of this step we found accurate fitting of body parts without requiring expensive training procedures. Finally, joint points are computed from the segmented body parts. The intersection of a thin plate orthogonal to the body crossing each joint point and the body hull selects which pixels will be used for measuring the corresponding trait.

To validate our work, we need a motion capture data set with limbs pixel labels and traits ground truth. Therefore, We have validated our proposed system on a novel data set of human poses, showing high segmentation accuracy and soft biometrics estimation. In particular, we found better segmentation performance than random forest approach.

The rest of the paper is organized as follows: Section 2 presents the details of our method, then experiments and results are described in Section 3, and finally we conclude our work in Section 4.

2. Limbs labeling and size measurements

In this section we review 3D shape context and TPS, describe our system for human limb segmentation and soft biometrics computation, whose different modules are shown in Fig. 2.

2.1. Training

The histogram of oriented gradients (HOG) descriptor has been studied vastly in the domain of human detection and pose recognition. Here, the key idea is to use HOG as the depth descriptor of the human body on depth images, where the gradients of the depth image are the derivatives of the body hull surface.

Once HOG feature vectors have been computed, our approach is based on modeling homogeneous pose clusters within a training set of depth human poses using a multi-class classifier. Then the sample pose models are computed from the nearest neighbors of each cluster. We use a problem-dependent clustering strategy to group HOG feature vectors of poses, as described next.

To cope with the problem of determining the exact number of clusters, we estimate the optimum number of clusters by combining the EM and k-means algorithms as proposed in [31]: let