Expert Systems With Applications 91 (2018) 270–276

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Enhanced automated body feature extraction from a 2D image using anthropomorphic measures for silhouette analysis

Simon Ouellet, François Michaud*

Interdisciplinary Institute for Technological Innovation (3IT), Université de Sherbrooke, 3000 Université Blvd, Sherbrooke, Québec, Canada

ARTICLE INFO

Article history:
Received 15 May 2016
Revised 20 July 2016
Accepted 4 September 2017
Available online 8 September 2017

MSC:
00-01
99-00

Keywords:
Body feature points
Human feature extraction
Anthropomorphic measures

ABSTRACT

Using mobile systems (e.g., smart phones, tablets, mobile robots), human metrology (HM) is a soft biometric that can be beneficial for human detection and identification from a 2D image, as long as it can be done online and in open conditions. This paper presents an approach that derives HM measurements from 2D images of silhouettes. The approach enhances an algorithm for automated body feature extraction from a 2D image in front of a black background, by using anthropomorphic information to extract, online and in parallel, 20 front and 13 side measures out of 45 front and 24 side features, increasing the number of features and improving processing time. Recognition and identification results are presented with both uniform (black) and real backgrounds, for distances ranging from 1 m to 6 m.

1. Introduction

Person identification from images is conducted for a great variety of applications, e.g., forensic context using surveillance cameras by looking at facial features, hands, body height or gait (Gibelli et al., 2016), monitor stadium entrance using near and wide field of view cameras (Kacperski, Wlodarczyk, Grabowski, & Neves, 2014), conduct human action recognition using conventional (Cheng, Liu, Wang, Li, & Zhu, 2015; Wu & Shao, 2013) or RGB-D cameras (Jalal, Kamal, & Kim, 2015). The application has a direct influence on the resolution, precision and quality of the data available to make the required analysis, which may involve human silhouette processing. Our application of interest is to be able to discriminate people from other objects using images taken from a mobile device, such as smart phones, tablets or mobile robots, equipped with a 2D camera. Face recognition (Mykoniatis, Angelopoulou, Schaefer, & Hancock, 2013; Zhang, Hornfeck, & Lee, 2013) and speaker identification (Grondin & Michaud, 2012; Ming, Hazen, Glass, & Reynolds, 2007) are biometrics that have been demonstrated on mobile robots, and using both can help improve robustness in unconstrained and dynamic environments (Stiefelhagen et al., 2007). Lawson and Martinson (2011) use a Markov Logic Network to fuse biometric indicators (nose, forehead and clothing images taken relative to the detected face position) with face recognition and speaker identification, demonstrating 74% recall and 97% precision, with people facing forward 1 m to 2 m away from the robot. However, these approaches require having people stand close and right in front of the device, making interaction restricted to a limited and constrained space, and activating data analysis once a person is placed in the field of view of the camera.

To enhance people identification from a distance and under different perspectives and poses, and making the system capable of detecting autonomously when a human is in its field of view, human metrology (HM) features (e.g., body shape, anthropometric measurements and geometrical features generated from these measurements) could be useful. HM can be defined using 10 measures of lengths (arm, head), breadths (head, shoulder), circumferences (armscye, chest, neck base, waist), stature and weight (Adjeroh, Cao, Piccirilli, & Ross, 2010). Circumference measures are difficult to extract from 2D images and require precise 3D modelling (Adjeroh et al., 2010; Godil & Ressler, 2011; Hilton et al., 2000). However, HM measures such as height (Madden & Piccardi, 2005) and silhouette (Burri, 2007) can be perceived from 2D images. Detecting a silhouette can then trigger the use of other modalities for people identification.

In 2011, Lin and Wang (2011) presented an interesting algorithm for body feature extraction using a 2D image taken from an inexpensive 2D camera. Using a higher number of features compared to six other methods, the algorithm extracts 38 front and 22 side

* Corresponding author.
E-mail addresses: Simon.Ouellet@usherbrooke.ca (S. Ouellet), Francois.Michaud@usherbrooke.ca (F. Michaud).

http://dx.doi.org/10.1016/j.eswa.2017.09.006
0957-4174/

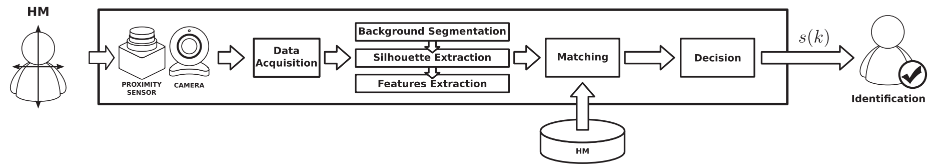

Fig. 1. HM recognition and identification approach.

features from a single silhouette in approximately 30 s at 3 m. The algorithm uses a clockwise sequential extraction process initiated from a starting point (the top of the head), to then follow the silhouette and extract the required features in a specific order. Validation was done for 30 subjects (15 males, 15 females) in front of a black background, with 100% feature recognition rate.

Lin and Wang's approach only works for silhouette detection in controlled conditions, and using this algorithm in real life settings on mobile devices requires two significant improvements: processing time must be reduced substantially, and the algorithm must be made more robust to noise in uncontrolled environments. This paper presents such improvements by integrating a dynamic background segmentation technique to support silhouette detection with realistic backgrounds, and by using a parallel feature extraction method based on an anthropometric search, making the algorithm fast enough for online processing and for minimizing recognition latency. Our approach also handles more features, i.e., 45 front and 24 right features, in real-time. To study how much can be done with the silhouettes extracted, we also evaluate if HM features can be used not only to recognize a silhouette, but also to identify the person.

The paper is organized as follows. Section 2 presents our approach and the improvements made to Lin and Wang's algorithm. Section 3 describes the experimental setup and configuration for the trials, along with the results.

2. Online human metrology recognition and identification from a 2D image

As shown by Fig. 1, our approach consists of six modules used in sequence: Data Acquisition of both camera and proximity sensor, Background Segmentation, Silhouette Extraction, Features Extraction, Matching and Decision. The Decision module provides an ordered list of k identification candidates from a set of K models, along with a confidence level $s(k)$.

2.1. Data acquisition

A Logitech c910 camera is used to capture 8-bit per channel RGB images at 10 Hz with 1280×960 pixel resolution (as in Lin & Wang, 2011) and standing at 1 m above the floor. A pinhole model is used to cancel lens distortions and to estimate the camera intrinsic parameters. Using the camera CCD/CMOS physical size, the horizontal and vertical field of view (FOV), the size of a pixel in mm can be calculated to enable conversion from pixel to mm.

Compared to Lin and Wang's algorithm, no image compression is done on the resulting silhouette extraction: only cropping of the region of interest is used. Also, images are converted from RGB to HSV color space to enhance background segmentation. In addition, a laser range finder is used to retrieve the distance between the system and the person for precise distance measurements.

2.2. Background segmentation

The first step for HM recognition and identification consists of removing the image background to extract the foreground silhouette, as illustrated by Fig. 4. This step directly influences the accuracy and precision of the foreground silhouette, and consequently the overall recognition and identification performance. With our approach, two background segmentation methods are used:

- For images of subjects placed in front of a black background and taken by a fixed camera, a binary threshold is used to classify pixels that belong to the silhouette. As done in Lin and Wang (2011), this condition is used to take images of people wearing white underwear in front of a black background and under controlled illumination conditions. This provides precise body measurements which would not be possible with people wearing casual clothing. Front and back images are taken with people adopting a standard posture with their limbs straight and arms apart from the torso, while left and right images are taken with their limbs straight and close to the body and feet close to each other.
- For images taken by the camera in a real environment, we integrated ViBe (Droogenbroeck & Paquot, 2012), a universal background substraction algorithm. ViBe is a pixel-based nonparametric background modelling method that represents background by a set of randomly-selected samples for each pixel. It compares new frames, pixel by pixel, to determine if a pixel belongs to the background or to the foreground. This approach works with possible occlusions caused by obstacles and in front of different types of static or changing backgrounds, an important requirement to use the approach on mobile devices: it enables the approach to detect motion when the device is immobile, and can then reinitialize the background model using only a single frame. ViBe takes a pre-processed image, models the background distribution at each pixel with randomly selected samples from pixel neighbourhood and a random update rule, and returns a binary image of the foreground. Note however that in real environments, such background segmentation provides coarse foreground contours which potentially may contain holes and undesired foreground objects. As a solution, our approach uses a blob detector to discriminate foreground regions present in the segmented image, as shown by Fig. 2. A blob is defined as a group of connected pixels that share similar information (Lindeberg, 1994). The largest region is labelled as the human silhouette while the others are labelled as foreground noise and are removed.

2.3. Silhouette extraction and representation

Similarly to Lin and Wang's algorithm, our approach uses a Canny edge detector (Canny, 1986) to extract the silhouette contour and to remove silhouette holes by filling the inner contour using closed morphology. The silhouette length can change based on orientation, distance and noise produced by human clothes and background segmentation. A Gaussian smoothing filter is applied to the coarse silhouette to reduce noise produced by background segmentation. Then, the silhouette contour is represented using Freeman eight-directional chain code (Freeman & Davis, 1977). As illustrated by Fig. 3, and starting from the top-left pixel of the silhouette, turning clockwise, chain code numbers from 0 to 7 are used to represent the direction of the silhouette by increments of 45°. The result is a closed silhouette curve represented