Journal of Applied Logic ••• (••••) •••–•••

Contents lists available at ScienceDirect

Journal of Applied Logic

www.elsevier.com/locate/jal

A brief review of the ear recognition process using deep neural networks

Pedro Luis Galdámez *, William Raveane, Angélica González Arrieta

University of Salamanca, Plaza de los Caídos, 37008 Salamanca, Spain

A R T I C L E   I N F O

Article history:
Available online xxxx

Keywords:
Ear recognition
Convolutional Neural Network
CNN

A B S T R A C T

The process of precisely recognize people by ears has been getting major attention in recent years. It represents an important step in the biometric research, especially as a complement to face recognition systems which have difficult in real conditions. This is due to the great variation in shapes, variable lighting conditions, and the changing profile shape which is a planar representation of a complex object. An ear recognition system involving a convolutional neural networks (CNN) is proposed to identify a person given an input image. The proposed method matches the performance of other traditional approaches when analyzed against clean photographs. However, the F1 metric of the results shows improvements in specificity of the recognition. We also present a technique for improving the speed of a CNN applied to large input images through the optimization of the sliding window approach.

1. Introduction

The study of biometrics is based on physical or behavioural characteristics of an individual in order to verify his or her identity. Features like fingerprints, face, and iris have received major attention for long time. Researchers consider fingerprints and iris are more precise in biometric analysis than the face, but the face has other qualities like being easily obtained in real conditions without user interaction. However the face by itself is not as flexible as it should be due to illumination and expression changes.

Ear images can be obtained with the same approach like the face, this scenario suggests that it could be used as a complement in a recognition system. Multiple researchers have affirmed that the ears are indeed unique enough to identify a person and they could have a practical use as a biometric feature [17].

2. Background

Fundamentally most of the ear detection approaches rely on properties in the ear's morphology, like the occurrence of certain characteristic edges or frequency patterns. Significant progress has been made in the

* Corresponding author.
E-mail address: peter.galdamez@usal.es (P.L. Galdámez).

http://dx.doi.org/10.1016/j.jal.2016.11.014

past few years in the ear biometrics field. One important technique known in ear detection was introduced by Burge and Burger [1]. They proposed a technique for detection using deformable contours, the main problem is that this method requires user interaction for contour initialization; therefore, the task of localization is not fully automatic. Hurley et al. [10] applied force fields, this process does not require to know the location of the ear to perform detection; however, this only applies in controlled environments without any kind of noise.

One of the most impressive techniques known to detect the ears is raised by A. Cummings et al. [8] who show a strategy using the image ray transform (IRT) which is capable of highlighting the ear tubular structures. The technique exploits the helix elliptical shape to calculate the localization. In [19], Yan and Bowyer have used manual technique based on two previous lines for detection, where takes a line along the border between the ear and face while another line crosses up and down the ear. In the context of 3D images, Zhou et al. [20] presented a novel shape based feature set, called Histograms of Categorized Shapes (HCS), for robust 3D ear detection. Using a sliding window approach and a linear Support Vector Machine (SVM) classifier.

Chen and Bhanu propose three different approaches for ear detection. In the first of these, they trained a classifier that recognizes a specific distribution of shape indices [4]; however, this only works on profile images and is very sensitive to any kind of rotation, scale or pose variation in the image. Later, they worked on image regions with a large local curvature using a technique known as step edge magnitude [3], where a template containing the typical shape of the outer helix and the anti-helix, is fitted to clusters of lines. Finally, they narrowed the number of possible ear candidates by detecting the skin region; first before applying the helix template matching to the curvature lines [5]. There are many proposals to solve the problem, this paper only has done a small review from some of them, in order to deepen about the literature review in the ear biometrics researches you can refer the work of Pflug et al. [15].

The Convolutional Neural Network (CNN) [7] has become a general solution for image recognition with variable input data. CNNs consist of two stages one for automated feature learning, and another for classification, both of them can be successfully trained in tandem through gradient descent of the error surface [14]. Its results have consistently outclassed other machine learning approaches in large scale image recognition tasks [11], outperforming even human inspection of extensive datasets [6].

Compared to other feature-based computer vision methods such as SIFT [13] or HOG [7], CNNs are much more robust and tolerant to shape and visual variations of the images or objects intended to be recognized. However, contrary to such methods, an execution of a CNN will only recognize features on a single image block of size equal to the input dimensions of the network. As CNNs are usually trained with small image patches, this recognition area is likewise small. As a result, to run image recognition over a larger image size, it is necessary to repeatedly apply the same network over multiple regions. This is a very common technique named sliding windows, albeit a time consuming one as the execution time naturally grows in proportion to the number of sampled blocks.

3. System description

The network on which our system is based upon is a standard CNN composed of alternating convolutional and max-pooling layers for the feature extraction stage, and one or more linear layers for the final classification stage. Fig. 1 depicts the layer structure of such a network, and it is the reference architecture used here to describe the concepts of the framework presented. The first layer in the network consists of one or more neurons containing the image data to be analyzed, usually composed of a single grayscale channel of the incoming image.

Recognition systems traditionally follow a set of standards, such as, acquiring images, pre-processing, feature extraction, classification, and/or recognition of the respective object. All of these tasks will be described in upcoming sections connecting important algorithms in order to complete its goal. Nevertheless,