Future Generation Computer Systems 88 (2018) 28–39

Contents lists available at ScienceDirect

Future Generation Computer Systems

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

Appearance based pedestrians' gender recognition by employing stacked auto encoders in deep learning

Mudassar Raza [a], Muhammad Sharif [a], Mussarat Yasmin [a,*], Muhammad Attique Khan [a], Tanzila Saba [b], Steven Lawrence Fernandes [c]

[a] Department of Computer Science, COMSATS Institute of Information Technology, Wah Cantt, Pakistan
[b] College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia
[c] Department of Electronics and Communication Engineering, Sahyadri College of Engineering & Management, Mangaluru, India

HIGHLIGHTS

- The pedestrians' gender is recognized based on body appearance.
- The stacked sparse auto encoder in deep learning is used for predictions.
- Performance is evaluated on the three views; front, back and mixed.
- Improved and acceptable outcomes are achieved in terms of accuracy.
- MIT and PETA datasets are used for experimentation and evaluation.

ARTICLE INFO

Article history:
Received 29 September 2017
Received in revised form 26 April 2018
Accepted 1 May 2018

Keywords:
Deep learning
Stacked sparse auto encoder
Pedestrians' gender classification
Pedestrian parsing
Soft max classifier

ABSTRACT

Pedestrians' gender is a soft attribute which is useful in many areas of computer vision including human robot interaction, intelligent surveillance and human behavior analysis. Apart from its importance, pedestrians' gender prediction is one of the challenging methodologies in image processing. In this article, a deep learning approach is presented to classify a pedestrian as a male or a female. As a pre-processing step, pedestrian parsing is performed by a deep decompositional neural network method. The outcome of this network is a binary mask that maps the pedestrian full body from the input image. The pedestrian body image is then extracted by applying the generated pedestrian mask to the input image. This pre-processed image is then supplied to the stacked sparse auto encoder with soft max classifier for prediction. The proposed network is trained and tested separately on different pedestrians' views such as frontal views, back views and mixed views. The training is performed on PETA dataset. The experiments for testing are performed on MIT and PETA datasets (containing images other than train images). The accuracy values on MIT dataset are calculated as 82.9%, 81.8% and 82.4% on frontal, back and mixed views respectively. The mean AUC value by proposed scheme on PETA dataset is found as 91.5%±4. The performance measures and comparisons with existing works depict the robustness and applicability of proposed methodology.

1. Introduction

In recent years, research is focused towards improving the automated friendly pedestrian environment. The intelligent transport system (ITS) methodologies are being used to answer pedestrian safety problem in streets. The problems like a pedestrian crossing the road and pedestrians having impairments tribulations have been the focus of ITS researchers [1]. In modern days, people are adopting security monitoring through cameras. Surveillance cameras are being seen everywhere from roads to buildings. Automated monitoring of big video data through camera surveillance is becoming a big demand to reduce human burden. Also, automated monitoring can be utilized to analyze pedestrians' behavior. Automated gender recognition has been a demanding field to research over past many years. In real, human beings differentiate people as male or female based on gait, voice, clothes, jewelry, hair styles, body shapes, face looks [2–9] and skin color. Tremendous approaches have been presented in this area based on facial

* Corresponding author.
E-mail addresses: mudassarkazmi@yahoo.com (M. Raza), muhammadsharifmalik@yahoo.com (M. Sharif), mussarat@ciitwah.edu.pk (M. Yasmin), attique@ciitwah.edu.pk (M.A. Khan), tsaba@psu.edu.sa (T. Saba), steven.ec@sahyadri.edu.in (S.L. Fernandes).

https://doi.org/10.1016/j.future.2018.05.002
0167-739X/

features [10–13]. The face-based gender recognition requires a proper visible face for prediction. While for the pedestrians' case, the images are usually acquired from the cameras that are at a far distance from the pedestrians. The distant images sometimes cause poor visibility of pedestrian's face. This is because the pedestrians objects are at a distant location from the camera and sometimes the resolution of face region becomes lesser than 16×16 [14]. Such type of images lie under very low resolution (VLR) image problems. The studies show that at least 32×32 resolution is required for face recognition [15]. In addition, the face-based recognitions also fail when the camera is located on the backside of person. The whole body and the pedestrian appearance can be a cue for predicting pedestrian's gender. The loose-fitting clothes can also create a problem in understanding exact body shape. The strong cues for such scenario can be body appearance that covers attributes like long/short coats, skirts, long/short hair, pants, shorts and shoes.

The automated recognition system that identifies humans as either male or female can be useful in a variety of applications. Some are: (a) Counting total number of men and women in a crowd [16] (b) Automated behavior analysis of males and females in a shopping mall (c) Human and robots communication where intelligent robots require gender information to perform appropriate communication with humans [17,18] and (d) Intelligent Transportation System to ensure pedestrians safety [1]. In this paper, the intention is to recognize pedestrians as male or female. Main contributions of this article include the following.

i. The gender is recognized based on body appearance.
ii. Pedestrian parsing procedure is followed which creates a silhouette of the input pedestrian image. That silhouette image is used as a binary mask to remove background from the input image.
iii. The stacked sparse auto encoder (SSAE) is used for pedestrians' gender prediction.
iv. The performance of gender is evaluated on different pedestrian orientations such as front, back and mixed views.
v. Improved and acceptable outcomes are achieved in terms of accuracy.

The remaining portion of paper is organized as follows: Section 2 highlights the contributions of some state-of-the art methods and found literature. Section 3 belongs to materials and methods, represents the proposed system. This section also highlights the architecture and configuration of the proposed SSAE. Section 4 presents the discussions about experiments and their results. Section 5 is a conclusion which is preceded by references. Table 1 shows mathematical representations that are employed in this manuscript.

2. Related work

In existing literature, different categories of gender recognition methods found are human face based [19–23], gait based [24–26], emotions based [27] and human body based methods [28–30]. Most of the work is performed on recognizing gender through face images. A very few research articles classify gender through human body appearance. The video surveillance cameras, in most of the cases, have low resolution and take images from a distance. In case of low resolution images, faces are very difficult to analyze. The human body attributes (dress, body, shape, and gait) on the other hand, are still useful even in low resolution scenarios. Factors affecting gender classification are similarity of body shape, clothing styles and hair styles of both males and females. In addition to these factors, some more challenges are pose variations, occlusion in images, age variations, very dim resolution of distant images and image noising.

For recognition, the two important tasks are feature extraction and classification [31–39]. The body-based gender recognition is a hard task because of huge variations in body shape. The problem is claimed to be addressed first by Cao et al. [40]. The researchers use a part-based algorithm which is applicable only on frontal and back views of the human body. The work is based on cue-based ensemble learning algorithm with AdaBoost and random forest (RF) as flexible classifiers. Unlike part based features, Matthew et al. [41] calculate features for the whole image. The authors use pyramid histogram of gradient (PHOG) as a shape descriptor and pyramid histogram of visual words (PHOW) as the appearance descriptor along with color histogram data for feature collection. These features are then sent to support vector machine (SVM) for classification. Furthermore, the authors also present their own dataset for gender recognition based on MIT dataset [42]. The dataset contains frontal images of 305 males and 123 females. The original version of the dataset is however not yet released. Matthew et al. [43], in their another research work, use linear discriminant analysis (LDA) through which they identified major principal component analysis (PCA) components for gender profiling. The focus is to analyze frontal human body shape from noisy pedestrian images. Guo et al. [16] propose manifold learning along with biologically inspired features for gender classification. The orientation classification is then performed as a next step. Linear support vector machines (L-SVMs) are further applied on resultant features for classification. The results are taken only on frontal and back views. Chen et al. [44] describe semantic clothing attributes for gender recognition. The upper portion of body is selected only for semantic attribute learning and gender prediction. Various clothing rules, usually followed by humans, are described using conditional random fields (CRF). The features selected are taken from scale invariant feature transform (SIFT) and then passed to SVM for classification. The AdaBoost algorithm is enhanced by Gou et al. [45] and used as a classifier for gender. Multiple scale texture (MST) feature is proposed by combining Haar and texture like features. Zhu et al. [46] propose methods for both single class and multi-class recognition of attributes. Histogram of oriented gradients (HOG) features with gentle AdaBoost algorithm are used for single class, while, K nearest neighbor (KNN) is used for multi-class classification. Ma et al. [47] use Denoeux's rule for combining multiple methods introduced over time. This is also known as Dempster Shafer (DS) evidential method.

Video based gender profiling is proposed by Wong et al. [48]. Back propagation neural networks are applied on images which are extracted from different movies by using background subtraction algorithm. Hall and Perona [29] use fine grained classification for pedestrians on video datasets.

Gender classification for a 3D view of upper body is introduced by Wang and Xia [49]. The silhouette images of the upper body are segmented first and then normalized elliptic Fourier descriptors are applied. PCA is used for dimensionality reduction and the resultant features are moved to back propagation neural network for classification. 3D body shape gender recognition is introduced by Luo and Wu [18]. Apart from the torso, shoulder width and chest features are selected and classified through SVM. 3D data from the side and back views are used with HOG and deep learning tessellation based approaches [28].

Human gender recognition based on Convolutional Neural Networks (CNN) has been the point of interest in recent years. Different variants of CNN architectures have been used to classify humans based on their faces. The researchers claim promising results on different available and standard face datasets [19,50–53]. C.-B. Ng et al. use 7 layers CNN architecture [54,55] for gender recognition and claim 80%–81.5% accuracies on frontal and rear views of full body pedestrians' images. The evaluation is performed on MIT pedestrian dataset which is a tiny dataset for deep learning