Pattern Recognition Letters 68 (2015) 297–305

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Shape Context for soft biometrics in person re-identification and database retrieval☆

Athira Nambiar*, Alexandre Bernardino, Jacinto Nascimento

Institute for Systems and Robotics, Instituto Superior Técnico, Av. Rovisco Pais, 1, Lisbon, 1049-001, Portugal

ARTICLE INFO

Article history:
Available online 11 July 2015

Keywords:
Multimedia
Soft biometrics
Shape Context
Re-Identification
Silhouettes
Retrieval
Surveillance

ABSTRACT

We introduce a novel descriptor for the analysis of pedestrians and its applications to person re-identification and database retrieval. A Shape Context descriptor of the head-torso region of persons' silhouettes is shown to have a very good discrimination ability and application to re-identification. For database retrieval using human queries, we train a map from the Shape Context to interpretable soft biometric quantities that can be reasoned about by humans. We show that a good linear correlation exists between Shape Context descriptors and soft biometrics quantities in the upper human torso and illustrate its application to retrieval in databases from human queries. Shape Context to biometrics maps are learned from virtual avatars rendered by computer graphics engines, to circumvent the need for time-consuming manual labelling of data sets. We obtained promising results of Shape Context based person re-identification and database retrieval from human compliant description of biometric traits, in both synthetic data and real imagery.

1. Introduction

Soft biometrics are the physical, behavioral or adhered human characteristics, classifiable in predefined human compliant categories which are established and time proven by humans with the aim of differentiating individuals [7]. Different from hard biometrics, they lack the distinctiveness and permanence to identify an individual with high reliability. However, they have certain advantages over hard biometrics exclusively making them best suited to deploy in surveillance applications viz. non obtrusiveness, acquisition from distance, non-requirement for the cooperation of the subject, computational and time efficiency and human compliance.

Soft biometric features leverage human characteristic traits such as height, body size and gait. These characteristics are more coherent and reliable for long term re-identification than the commonly used temporary appearance cues such as dress color and texture info [5]. Recently, the arrival of sophisticated systems such as motion capturing devices, 3D sensors (Kinect) and high definition cameras accelerated the exploitation of soft biometrics in wide range. As a result, unprecedented real time applications were reported in person re-identification and other video surveillance applications.

However the direct computation of soft biometric features from video images is not trivial and existing methods rely on human manual measurements made on individual images [17]. Instead, automated computer vision analysis methods have been more successful with features that are not interpretable by humans, like SIFT [13], HOG [6], Shape Context [4] and others. These features, though useful in automated methods, are hard to reason about by humans and thus not suited for formulating verbal descriptions of search queries in databases. For instance, we would like to be able to search on a database for persons with large torso, thin neck, long head, etc. Thus, we propose a methodology to infer soft biometric person characteristics from their computer vision based descriptors, using regression analysis.

Obtaining a predictive model of soft biometric features from computer vision features involves several challenges and difficulties: (i) which computer vision features are more adequate; (ii) how to obtain the ground truth biometric features to train the model and; (iii) which regression model is more suitable.

With respect to the first point, we propose the use of Shape Context features computed in the upper-torso part of the frontal human silhouettes, where we capture the human images from their video clips walking towards the camera. The upper torso region of the body presents less temporal variance with respect to arms and legs motions, thus producing more stable features. In addition to that, since person re-identification is carried out in an uncontrolled environment, there are chances for clutters and other interacting objects making the lower body part occluded. In many indoor surveillance systems cameras are placed along corridors at high positions and tilted down, which makes the legs and lower torso occluded when persons are close to the camera. However, the head to chest

☆ This paper has been recommended for acceptance by Paulo Lobato Correia and Thomas Moeslund.

* Corresponding author. Tel.: +351 218418050.
E-mail address: anambiar@isr.ist.utl.pt, nambiar.athira@gmail.com (A. Nambiar).

http://dx.doi.org/10.1016/j.patrec.2015.07.001
0167-8655/

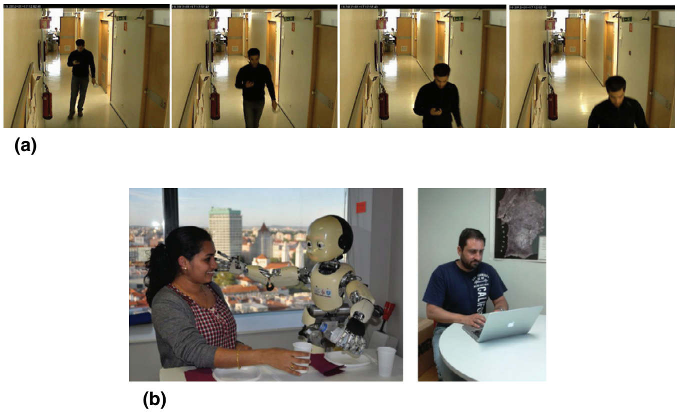

Fig. 1. Images showing the relevance of head-to-chest region for person re-identification tasks in various situations. (a) A forward walking sequence captured in our HDA data set [14] highlights the visibility of head-to-chest region in most of the frames, while the other parts are occluded. (b) Relevance of Upper body region in Human-Robot/Computer interaction. Head to chest region are clearly visible in both cases and could be used for re-identification task.

region, unlike the waist and legs, maintain a relatively consistent shape through a broader range of walking frames. A real world example (video sequence in the HDA Data set[1]) where this effect is clear is shown in Fig. 1(a). In addition to that, there are scenarios in Human-Robot Interaction (Fig. 1(b) left) and Human-Computer Interaction (Fig. 1(b) right) that highlight the relevance of upper body part selection for person re-identification.

The use of a silhouette based feature is motivated by the fact that it is less sensitive to the color and texture of the inner region of person's images and thus making itself a better candidate towards long term based person re-identification. Furthermore, the Shape Context feature computes the density of boundary points at various distances and angles. As such, it more directly encodes soft biometric traits such as lengths, curvatures and size ratios in the human body. In this paper, we explore this idea with the goal of recovering the soft biometric features encrypted in the Shape Context descriptors of body silhouettes using regression methods.

The second challenge is the availability of ground truth biometric features to train the model in the regression analysis. It is not easy to model this in a real environment due to the necessity of a range of variations of discriminative biometric features in relatively large population. Also, it is laborious to annotate the human biometrics manually on real data. In order to tackle this issue, we used synthetic avatars in a virtual reality platform. In contrast to [17], where the training set was generated by manual annotations done on real imagery by a large number of human annotators, here we avoid such a troublesome training phase by generating the ground truth with the help of modern computer graphics technology.

In this work we leverage on the ability to simulate thousands of variations in biometrics on avatars according to our choice for two purposes. First, we conduct a baseline study to verify the impact of our descriptor for re-identification (Re-ID), since the simulated avatars provide flawless silhouette images. Second, we model the regression between computer vision based features (Shape Context) and human interpretable features (Biometrics) and thus bridge the gap between the human and machine interpretations of human body shape. Thus we present a novel automatic person retrieval system which could work in dual mode (viz., multimedia mode or human query mode) depending on the test data.

For obtaining some geometric features of the head to chest region, distinguishable from person to person, some measurable metrics which vary significantly within the population should be chosen. The measurement and study of such features and their variation is the domain of anthropometry. An anthropometric survey (ANSUR) was conducted by the U.S. military in 1988 upon more than 150 anthropometric dimensions, measured from 9,000 soldiers. A statistical summary of those standard biometric features related to the upper torso regions is provided in Table 1. In our study, we consider some of those key biometric features described here.

Table 1
A summary of anthropometric data taken from [9] relevant in the upper torso region. These statistical summaries reveals significant variation in the head and chest measures (All measurements are in centimeters.) The features shown in bold letters are some of the soft biometric cues used in our study.

Measurement name	Mean	Standard deviation	Min	Max
Biacromial breadth	39.70	1.80	33.0	45.10
Bideltoid breadth	49.18	2.59	41.0	59.3
Head width	15.51	0.60	13.6	17.7
Head circumference	56.77	1.54	51.4	62.7
Head length	20.02	0.72	17.6	22.6
Chest breadth	32.15	2.55	25.70	42.20
Neck-bustpoint length	27.24	1.81	22.2	34.2
Neck circumference	37.96	1.97	31.6	47.0
Shoulder circumference	117.52	6.04	96.6	142.4
Shoulder-elbow length	36.9	1.79	29.7	44.6
Shoulder length	15.05	1.10	11.4	18.5

Concerning the regression model, several choices are possible to be used, e.g., a non-linear mapping based on Gaussian process regression [16]. However, this kind of approaches usually require a significant amount of data. In our case, we will adopt a linear approach as a baseline, to perform a mapping between the Shape Context and soft biometrics. We will experimentally demonstrate that with the available amount of data such linear approach will suffice for obtaining high accuracy in the results.

The paper is organized as follows. Section 2 describes related literature on soft biometric based person retrieval as well as former applications of Shape Context in surveillance scenarios. The system architecture is explained in detail in Section 3. In Section 4, the main methodologies used in our framework viz. Shape Context feature extraction and regression are described. Then, Section 5 explains in detail our Re-ID experiments conducted in both real and virtual platforms, by acquiring Re-ID data set and by simulating avatars respectively. The regression analysis carried out is also explained here. Section 6 presents a set of promising results accentuating the reliability of the proposed architecture in person re-identification, as well as demonstrating the human retrieval performance with human compliant queries. Finally in Section 7, we summarize our work and enumerate some future work plans.

2. Related work

2.1. Shape Context on surveillance applications

The application of Shape Context *(SC)* in human video surveillance systems are reported in the state-of-the-art. Some works are found in pedestrian detection by [12], highlighting that *SC* descriptor trained on real edge images exhibited high performance, particularly on difficult images and backgrounds. Some application of *SC* have also been employed in gait recognition [22] where *SC* is used to compute the similarity between two procrustes mean shape, which is a compact representation of gait sequence. A similar application of *SC* is found in human pose estimation [1]. However, the literature is scarce concerning the use of *SC* in re-identification applications. One exception is [21] that created shape labelled images by means of shape and appearance models which was inspired from the idea of Shape Context.

[1] http://vislab.isr.ist.utl.pt/hda-dataset/ .