Neurocomputing 100 (2013) 197–205

Contents lists available at SciVerse ScienceDirect

Neurocomputing

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

Demographic classification from face videos using manifold learning

Abdenour Hadid *, Matti Pietikäinen

Machine Vision Group, University of Oulu, P.O. Box 4500, FI-90014, Finland

ARTICLE INFO

Available online 1 June 2012

Keywords:
Manifold learning
Locally linear embedding
Gender recognition
Age classification
Local binary patterns
Spatiotemporal analysis
Demographic classification

ABSTRACT

Research on automatic demographic classification is still in its infancy despite the vast potential applications. The few existing works are only based on static images while nowadays input data in many real-world applications consist of video sequences. From these observations and also inspired by studies in neuroscience emphasizing manifold ways of visual perception, we propose in this work a novel approach to demographic classification from video sequences which encodes and exploits the correlation between the face images through manifold learning. Our extensive experiments on the gender and age classification problems show that the proposed manifold learning based approach yields in excellent results outperforming those of traditional static image based methods. Furthermore, to gain insight into the proposed approach, we also investigate an LBP (local binary patterns) based spatiotemporal method as a baseline system for combining spatial and temporal information to demographic classification from videos.

1. Introduction

Automatic demographic classification from human faces generally includes gender recognition (i.e. man vs. woman), age categorization (e.g. child, youth, adult, middle-age and elderly) and ethnicity classification (e.g. Asian, Caucasian and African). This is very useful for more affective human–computer interaction (HCI) and smart environments in which the systems should adapt to the users whose behaviors and preferences are not only different at different ages but also specific to a given ethnic and/or gender. Automatic demographic classification is also useful in many other applications such as content-based image and video retrieval, restricting access to certain areas based on gender and/or age, enhancing the performance of biometric identification systems, collecting demographic information in public places, counting the number of women entering a retail store and so on.

Though there has been a great deal of progress in face analysis in the last years, demographic classification tasks have not been associated to that progress as most work has mainly focused on face detection and recognition problems. Consequently, the design of algorithms that are effective in discriminating between males and females, or classifying faces into different age and ethnic categories is still challenging and remains an open area of research.

First attempts of using computer vision based techniques to gender classification started in early 1990s. Since then, a significant progress has been made and several approaches have been reported in literature. Fundamentally, the proposed techniques differ in (i) the choice of the facial representation, ranging from the use of simple raw pixels to more complex features such as Gabor responses and (ii) the design of the classifier, ranging from the use of nearest neighbor (NN) and Fisher linear discriminant (FLD) classifiers to artificial neural networks (ANN), support vector machines (SVM) and boosting schemes. For instance, Moghaddam and Yang [1] used raw pixels as inputs to SVMs while Baluja and Rowley [2] adopted AdaBoost to combine weak classifiers, constructed using simple pixel comparisons, into single strong classifier. Both systems showed good classification rates. A comparative analysis on gender classification approaches can be found in [3].

While gender recognition has been explored by many other researchers [3], automatic age and ethnicity classification problems have received relatively far less attention despite the vast potential applications. Among the notable attempts are the works of Lanitis et al. [4] and Geng et al. [5]. Lanitis et al. used a simple quadratic aging function to model the relation between face and age, while Geng et al. modeled the sequence of a particular individual's face images sorted in time order by a subspace in which unseen faces are then projected for age estimation. A recent survey on different methods for age estimation can be found in [6].

Recently, the local binary pattern (LBP) features [7,8] have been successfully applied to demographic classification from static images (e.g. by Yang and Ai [9]) and to age estimation (e.g. by Chen et al. [10]). More recently, the combination of global and local features has also been shown to provide very good results in age estimation [11].

* Corresponding author.
E-mail address: hadid@ee.oulu.fi (A. Hadid).
URL: http://www.cse.oulu.fi/MVG/ (A. Hadid).

0925-2312/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2011.10.040

1.1. Motivations

It appears that most proposed approaches to demographic classification, including those cited above, are based only on static images and assume well aligned faces while nowadays the input data in real-world applications (such as in video surveillance and HCI) generally consist of video sequences and it is not always obvious to hold the face alignment assumption. So, the question which arises then is how to efficiently perform automatic demographic classification from face video sequences? A straightforward approach would be applying methods developed for still images to some selected frames and then fusing the results at decision or score levels. Obviously, such an approach is not optimal as it only exploits the abundance of frames in the videos and ignores the temporal correlation between the face images. Only recently have researchers started to also pay an increasing attention to the facial temporal information especially for face and facial expression recognition from videos (e.g. [12–18]) using for instance, spatiotemporal representations. However, demographic classification tasks have not yet been investigated from such points of view. To the best of our knowledge, no previous work has yet even addressed demographic classification from face video sequences.

Inspired by studies in neuroscience emphasizing manifold ways of visual perception [19] and also motivated by the psychophysical findings (e.g. [20,21]) which indicate that facial temporal changes can provide valuable information to face and gender recognition, we consider in this work the problem of demographic classification from video sequences and propose a novel approach which exploits and encodes the correlation between the face images through manifold learning. Thus, we look at the problem of demographic classification from totally new perspectives.

The goal of face manifold learning is to discover the hidden low-dimensional structure of the face images. Thus, instead of treating each facial image as a "single" or "isolated" pattern in the image space and then fusing the results, we propose to learn and discover the hidden low-dimensional nonlinear manifolds of the faces in each demographic class (e.g. male class, female class, child class, etc.). In other terms, we cluster the face sequences in the low-dimensional space based on their intrinsic demographic characteristic. Then, a target face sequence can be projected into the manifolds for classification. The "closest" manifold, in terms of a newly introduced manifold distance measure, will then determine the gender (or age or ethnicity) of the person in the target sequence.

Recently, there has also been an increasing interest on face manifold learning but most of the works were devoted to face recognition with an aim of coping with pose and illumination changes in the videos (e.g. [22–24]). To gain insight into our proposed approach, we also derive and investigate a baseline system that uses an LBP based spatiotemporal representation for combining facial structure (i.e. spatial information) and dynamics (i.e. temporal information), and support vector machines (SVM) for classification. Our choice of adopting the LBP spatiotemporal representation is motivated by the recent success of using it for combining appearance and motion for face and facial expression recognition [18,17] and also for dynamic texture recognition [17].

1.2. Salient contributions

The preliminary results of the research on the gender recognition problem have been published in part as a conference paper in [25]. In this article, we include new experiments on two other demographic classification problems which are age estimation and ethnicty classification, and report the complete and improved formulation, thorough investigation and extended experimental evaluation of our methodology.

Among the salient contributions of this article are: (i) a novel manifold based method to gender and age classification from face sequences is presented and extensively evaluated; (ii) an extension to the locally linear embedding algorithm [26] to handle face sequences is proposed; (iii) a simple yet efficient manifold to manifold distance measure is introduced; and (iv) a comparison between still image and video based analysis for demographic classification is provided.

The rest of this paper is organized as follows. Section 2 describes our proposed approach to demographic classification from videos using manifold learning. Section 3 presents a baseline method using an LBP based spatiotemporal representation and SVMs. Then, we present in Sections 4 and 5 the extensive experiments on two demographic classification tasks namely gender and age classification from videos. Section 6 further discusses the results and presents preliminary experiments on ethnicity classification. Finally, we draw a conclusion in Section 7.

2. Proposed approach

We describe below our proposed approach to demographic classification from videos. For clarity, we explain our derivations on the gender recognition case. Later, we show how to also apply the described methodology to the age estimation problem.

The idea of approaching demographic classification from manifold learning perspective is inspired by neuroscience studies pointing out the manifold ways of visual perception [19]. Indeed, the facial images are not "isolated" patterns in the image space but lie on a nonlinear low-dimensional manifold. The first key issue is to learn the face manifolds and discover the shared characteristics of the faces in each demographic class. This can be done by embedding the face images into low-dimensional coordinates. For that purpose, there exist several methods. The traditional ones are Principal Component Analysis (PCA) and Multidimensional Scaling (MDS). These methods are simple to implement and efficient in discovering the structure of data lying on or near linear subspaces of the high-dimensional input space. However, face images do not satisfy this constraint as they lie on a complex nonlinear and nonconvex manifold in the high-dimensional space. Therefore, such linear methods generally fail to discover the real structure of the face images in the low-dimensional space. As an alternative to PCA and MDS, one can consider some nonlinear dimensionality reduction methods such as Self-Organizing Maps (SOM) [27], talent variable models [28], Generative Topographic Mapping (GTM) [29], Sammon's Mappings (SM) [30], etc. Though these methods can also handle nonlinear manifolds, most of them tend to involve several free parameters such as learning rates and convergence criteria. In addition, most of these methods do not have an obvious guarantee of convergence to the global optimum. Fortunately, in the recent years, a set of new manifold learning algorithms have emerged. These methods are based on an Eigen decomposition and combine the major algorithmic features of PCA and MDS (computational efficiency, global optimality, and flexible asymptotic convergence guarantees) with flexibility to learn a broad class of non-linear manifolds. Among these algorithms are locally linear embedding (LLE) [26], ISOmetric feature MAPping (ISOMAP) [31] and Laplacian Eigenmaps [32].

We adopt in our work the LLE approach for its demonstrated simplicity and efficiency to recover meaningful low-dimensional structures hidden in complex and high dimensional data such as face images. LLE is an unsupervised learning algorithm which