Expert Systems with Applications 42 (2015) 269–279

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Sparse multi-stage regularized feature learning for robust face recognition

Mohamed Anouar Borgi [a,*], Demetrio Labate [b], Maher El Arbi [a], Chokri Ben Amar [a]

[a] Research Groups on Intelligent Machines, University of Sfax, BP 1173, Sfax 3038, Tunisia
[b] Department of Mathematics, University of Houston, Houston, TX 77204, USA

ARTICLE INFO

Article history:
Available online 6 August 2014

Keywords:
Neural networks
Shearlets
Sparsity
Shearlet Networks
Wavelet Networks
Face recognition

ABSTRACT

The major limitation in current facial recognition systems is that they do not perform very well in uncontrolled environments, that is, when faces present variations in pose, illumination, facial expressions and environment. This is a serious obstacle in applications such as law enforcement and surveillance systems. To address this limitation, in this paper we introduce an improved approach to ensure robust face recognition, that relies on two innovative ideas. First, we apply a new multiscale directional framework, called Shearlet Network (SN), to extract facial features. The advantage of this approach comes from the highly sparse representation properties of the shearlet framework that is especially designed to robustly extract the fundamental geometric content of an image. Second, we apply a refinement of the Multi-Task Sparse Learning (MTSL) framework to exploit the relationships among multiple shared tasks generated by changing the regularization parameter during the recognition stage. We provide extensive numerical tests to show that our Sparse Multi-Regularized Shearlet Network (SMRSN) algorithm performs very competitively when compared against different state-of-the-art methods on different experimental protocols, including face recognition in uncontrolled conditions and single-sample-per-person.

1. Introduction

Face recognition (FR) is one of the main areas of investigation in biometrics and computer vision. It has a wide range of applications, including access control, information security, law enforcement and surveillance systems. Despite the intense research activity in this area, however, many fundamental challenges remain and "at their current level of development, current facial recognition systems show promise but are not yet advanced enough to be considered mature technologies", according to a statement recently found on the face recognition page of the RAND Corporation (http://www.rand.org/natsec_area/products/facialrecog.html). In fact, current FR algorithms perform rather well under controlled conditions (i.e., with each subject looking directly into the camera, with good illumination, and filling the area of the photo completely) but the recognition rates suffer dramatically in the presence of variations in pose, illumination and facial expression. This is in contrast with the remarkable human ability for recognizing faces: an infant innately responds to face shapes at birth and can discriminate his/her mother's face from a stranger's within about 2 days. Thus, there is a need to develop more advanced face recognition methodologies that are capable of providing accurate recognition under realistic conditions. This is crucially important for all applications mentioned above.

The main objective of the ongoing research in FR is to improve the robustness of the algorithms with respect to the various factors indicated above and many of the methods proposed during the last decade have been trying to take advantage of the latest advances in statistical learning. In particular, an important class of recent FR methods focuses on the notion of *sparsity* to extract the salient facial features in such a way to obtain robust classification with respect to a wide range of changes in pose, illumination, etc. In nutshell, the philosophy underlying sparsity is that the essential information embedded in most multidimensional phenomena is intrinsically low dimensional, that is, there exists a sparse representation for such phenomena (Donoho, Vetterli, DeVore, & Daubechies 1998). The human ability to recognize faces can be interpreted as a manifestation of our ability to sparsely represent perceptual data or, in other words, to efficiently reduce the dimensionality of the data.

One of the main novelties of the method proposed in this paper is the use of the shearlet representation to extract the essential geometric content of facial features. This approach, pioneered by

* Corresponding author.
E-mail addresses: anoir.borgi@ieee.org (M.A. Borgi), dlabate@math.uh.edu (D. Labate), maher.elarbi@ieee.org (M. El Arbi), chokri.benamar@ieee.org (C. Ben Amar).

http://dx.doi.org/10.1016/j.eswa.2014.07.044
0957-4174/

one of the authors, is a powerful extension of the classical wavelet framework that combines multiscale analysis with directional sensitivity. Together with the curvelet approach (that shares some of its properties) the shearlet representation is the only method known to provide optimally sparse representations for images with edges (Guo & Labate, 2007), so that it can capture most efficiently edges and other landmark in images. These properties go far beyond the basic capabilities of PCA and their variants as well as conventional multiscale methods.

The other main novelty of our approach is the way we use the features extracted to learn from the data. The method that we use is a refinement of multi-task learning MTL, originally proposed by Caruana (1997). MTL attempts to learn classifiers for multiple tasks jointly and works under the assumption that all tasks should share some common features. In our algorithm, we use this idea to exploit the relationships among multiple shared stages of our recognition algorithm in such a way to obtain more robust recognition.

Thus, by combining the power of the sparse shearlet representation together with our refined version of multi-task learning, we introduce an improved framework for robust face recognition that we call Sparse Multi-Regularized Shearlet Network (SMRSN). This method includes a multi-regularization step inspired from multi-stage convex relaxation (Zhang, 2010) to upgrade from a non-convex optimization to a convex relaxation. As part of this work, we have extensively tested the performance of our FR algorithm under challenging uncontrolled conditions (i.e., changes of pose and illumination, and single-subject-per-person) and compared it against other standard and state-of-the-art algorithms. As we will show below, the performance of our method is outstanding.

The rest of this paper is organized as follows. In Section 2, we describe the related work on regularization theory and MTL. In Section 3, we provide the necessary background on shearlets. In Section 4, we present the proposed Sparse Multi-Regularized Shearlet Network (SMRSN) algorithm. In Section 5, we present extensive numerical experiments to demonstrate the efficacy of the proposed algorithm and compare it against standard and state-of-the-art methods. In Section 6, we have concluding remarks.

2. Related work

Face recognition has been an active field of research for more than two decades and many approaches have been proposed. Among the most classical and representative methods, we recall Eigenface (Turk & Pentland, 1991), Fisherface (Belhumeur, Hespanha, & Kriengman, 1997) and SVM (Heisele, Ho, & Poggio, 2001). As we mentioned above, the focus of current research is the development of improved FR algorithms that are highly efficient also in *uncontrolled* environments. One of the most successful contributions appeared in recent years is the algorithm proposed by Wright, Yang, Ganesh, Sastry, and Ma (2009), that applies a sparse coding technique to FR called Sparse Representation-based Classification (SRC). By coding a test (or query) image as a sparse linear combination of all the training samples, SRC classifies the image by evaluating which class could result in the minimal reconstruction error. One of the remarkable properties of this approach is its robustness to face occlusion and corruption which is due in part to the sparsity constraint applied on the coding coefficients. Following this idea, several variants of SRC have been proposed, such as the Regularized Robust Coding (RRC) by Yang, Zhang, Yang, and Zhang (2011) and Yang, Zhang, Yang, and Zhang (2013), that aims to reduce the computational cost of SRC and improve robustness. Its main idea consists in using a linear regression approach with regularized regression coefficients where, by assuming that the coding residual and the coding coefficient are respectively independent and identically distributed, RRC seeks for a maximum a posterior solution of the coding problem. The approach that we adopt in this paper follows the general idea of RRC. Also in our method, that we will describe in the Section 4, after encoding the image information using the shearlet framework, we will represent faces as sparse linear combinations of the training faces to achieve a classification. In addition to taking advantage of shearlets to efficiently encode the geometric information of images, the main difference of our approach with respect to RRC, is that we use a more sophisticated regularization approach for classification, which is inspired by ideas from Multi-Task-Learning.

Recall that, among the classification methods recently remerged in the machine learning literature, Multi-Task Learning (MTL) has been especially successful due to its remarkable performance. MTL aims to learn classifiers for multiple tasks jointly and works under the assumption that different tasks should share some common features. Many variants of MTL were proposed, including the multi-stage multi-task feature learning (MSMTFL) recently introduced by (Gong, Ye, and Zhang (2013)), who defines a non-convex formulation for multi-task feature learning based on a novel non-convex regularization, called *capped-* ℓ_1, ℓ_1 *regularized model* for multi-task feature learning. This approach aims to simultaneously learn the features specific to each task as well as the common features shared among tasks. A variant of this approach was proposed by Zhang (2010) and Zhang (2012) whose main novelty is a multi-stage convex relaxation scheme for solving problems with non-convex objective functions. Despite its success, very few publications have attempted to study the FR problem using MTL. One notable contribution in this direction is the work of Wang, Zhang, and Zhang (2009), that presents a modified multi-task learning (MTL) framework, called boosted MTL, for face verification with limited training data. This algorithm learns classifiers for multiple persons by sharing a few boosting classifiers in order to avoid overfitting. In our algorithm, we will adopt a refined version of MTL that uses a multi-task regularization to update the residual at every stage and a more effective approach to share classifiers. This improvement will be discussed in detail in Section 4.

3. Shearlet

The shearlet transform, introduced by one of the authors and his collaborators (Labate, Lim, Kutyniok, & Weiss, 2005), is a multiscale framework for image analysis that is especially designed to represent information not only across several scales and locations, as the conventional wavelet transform, but also across several orientations, in such a way that it can more efficiently encode geometric features such edges and other landmarks in images. The analyzing shearlet functions are scaled using a combination of shear matrices B_s, $s \in \mathbb{R}$ and anisotropic dilation matrices A_a, $a > 0$, defined by:

$$B_s = \begin{pmatrix} 1 & -s \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad A_a = \begin{pmatrix} a & 0 \\ 0 & \sqrt{a} \end{pmatrix} \tag{1}$$

It follows that the shearlet filters are highly anisotropic and can be oriented along any directions as illustrated in Fig. 1. Thanks to these properties, the shearlet transform has the ability to detect very efficiently the geometry of edges and other elongated features that usually constitute the dominating landmarks in many types of images, most notably face images. Related to this, shearlets provide optimally sparse approximation properties for the class of cartoon-like images, a simplified model of natural images, outperforming conventional wavelets and other traditional methods (Guo & Labate, 2007).