Neurocomputing 196 (2016) 70–81

Contents lists available at ScienceDirect

Neurocomputing

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

Sparse recognition via intra-class dictionary learning using visual saliency information

Ji-xin Liu [a,*], Quan-sen Sun [b]

[a] Engineering Research Center of Wideband Wireless Communication Technology, Ministry of Education, Nanjing University of Posts and Telecommunications, Nanjing 210003, PR China
[b] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, PR China

ARTICLE INFO

Article history:
Received 25 July 2015
Received in revised form
11 December 2015
Accepted 23 January 2016
Available online 17 March 2016

Keywords:
Sparse recognition
Dictionary learning
Visual saliency information
l_1-norm optimization

ABSTRACT

In recent years, sparse recognition (SR) has increasingly become an emerging pattern recognition method. Because of its excellent recognition performance for some traditionally difficult problems (such as occluded or corrupted face recognition), several classical SR ideas (such as sparse representation-based classification (SRC) or dictionary-based sparse recognition (DSR)) have been the focus of research in the intelligent information field. However, for image recognition against actual backgrounds, there are still problems with these mainstream SR methods. Hence, this paper presents a new SR method which combines the advantages of both SRC and DSR. In the pre-processing, visual saliency information (VSI) for images with complex scenes is extracted by introducing the saliency map as a tool. Then, DSR is used to develop intra-class dictionary learning for the VSI data. The last step is to solve a l_1-norm optimization problem to give the SR result by generating a global recognition matrix with the SRC mechanism. Experimental results show that the proposed method for 'real world' image recognition provides advantages over mainstream SR methods, in recognition rate and computation time cost.

1. Introduction

Sparse recognition (SR) as an extension of compressed sensing (CS) theory has been receiving more and more attention recently. This SR theory differs from traditional pattern recognition methods in that it skillfully utilizes l_1-norm optimization to realize a measure of generalized dependence between one test sample and its corresponding training sample set. In recent years, research on SR has generated plenty of academic achievements in many areas, such as face recognition [1–5], speech recognition [6–9], human behaviour recognition [10,11], text or number recognition [12,13], complicated target recognition [14–16], and so on.

Since SR possesses many advantages, it has become the trend and direction for current development and study in image recognition [17]. However, unlike the ideal conditions in a laboratory, imaging in real situations usually involves a lot of complicating factors (including unpredictable background, variable illumination, or uncertain positioning). Although using the SR method can achieve a satisfactory recognition rate for some experimental data sets [18], the results can become weak and unreliable in real conditions. Let us take some typical examples from face recognition. In the AR-related study [1,18], SR was able to achieve a recognition rate of 90% or more when all image samples with a flat background were aligned. In stark contrast, the recognition rate for SR struggled to reach 30% for data sets (such as labelled faces in the wild (LFW) [19]) from the real world. Thus it can be seen that the traditional SR method faces challenges in practical applications.

To solve the problems above, a method to optimize the generation of a recognition matrix (equivalent to the measurement matrix in CS framework) is widely thought to have the most promise. To this end, there are three typical methods. (1) Generating a global recognition matrix directly from the original image training set with alignment [20,21]. The advantage of this method lies in making the data much easier to deal with; the weakness is that the recognition rate is highly vulnerable to imaging condition changes. (2) Generating an exclusive recognition matrix for each class by the dictionary learning of each image training set [22,23]. This method will have a more robust effect on the imaging conditions, but the recognition result may trap into local optimum easily when the background is complicated. (3) Generating a global recognition matrix through the dictionary learning, with interest points in each image training sample (instead of the corresponding image) [24]. This pre-processing can mitigate the negative effects of image background; however, noise will be introduced if the selection of interest points is incorrectly

* Corresponding author.
E-mail address: liujixin@njupt.edu.cn (J.-x. Liu).

http://dx.doi.org/10.1016/j.neucom.2016.01.092
0925-2312/

made. In sum, finding a way to distinguish the key information from a complex image scene is the core task for image object recognition in practical applications. For the SR method it is not only a necessary process for recognition matrix optimization, but also the guarantee for recognition performance enhancement.

In this paper, we propose a new SR method based on intra-class dictionary learning using visual saliency information (VSI). Fig. 1 illustrates the basic processing flow of our SR method. We make the following two contributions.

Firstly, the VSI is extracted by using a saliency map for image recognition under natural conditions. Currently, how to extract the object key information from a complex image scene is a big problem. In this study, our solution is to introduce a saliency map for information extraction. A saliency map is a tool for the simulation of human visual perception [25]. It provides a conceptual framework that accounts for how the significant object of attention is selected by humans in natural situations. Because of this property, it has achieved great success in image processing (such as image segmentation or target detection [26–30]). It has also received some attention from the research into pattern recognition [31–34]. So this paper will improve the key information description under natural imaging conditions by using VSI instead of the original image.

Secondly, the global recognition matrix consists of the intra-class dictionary learning of each class image training set. The concept of the recognition matrix in SR arises from the under-determined random measurement in CS theory. Therefore, the most typical SR method – that is, sparse representation based classification (SRC) [1] – is to generate the global recognition matrix by using the original data directly. But considering the limitation of alignment pre-processing, dictionary-based sparse recognition (DSR) [22], which is different to SRC, selects the dictionary learning for each classification to generate individual recognition matrices. In other words, the l_1-norm optimization tactic in DSR is split into each classification for measuring the generalized pertinence between the test sample and the intra-class dictionary learning result of their corresponding training sample sets. Unfortunately, because of the lack of a global information measure, this method could easily fall into local optimization. To improve the SR performance, we adopt the SRC framework to generate a global recognition matrix which comprises the intra-class dictionary learning result of each class VSI set.

This paper is organized as follows: Section 2 gives an overview of prior work on the SR method. In Section 3 the proposed SR method via intra-class dictionary learning using VSI is detailed. Experimental results are presented in Section 4, and Section 5 concludes the paper with a brief summary and discussion.

2. Related work

2.1. Sparse representation-based classification (SRC)

It is generally known that the most representative SR method is SRC. The theoretical basis of SRC is primarily from CS theory. This theory was presented and proved by Candès et al. [35,36] and Donoho [37]. In CS theory, assuming the CS measurement matrix $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ and sensing data $y \in \mathbb{R}^M$ for any sparse signal $x \in \mathbb{R}^N$ have been acquired, this signal can be recovered by solving an l_1-norm optimization problem, as follows:

$$\hat{x} = \arg\min_x \|x\|_1 \ s.t. \ \mathbf{\Phi} x = y \tag{1}$$

Inspired by CS theory, Wright et al. [1] applied this idea to face recognition, and this method is known as sparse representation-based classification (SRC). Under the SRC framework, sensing data y is regarded as one test sample and the measurement matrix is converted to the recognition matrix, which consists of the training sample set

$$A = [A_1, \ldots, A_i, \ldots, A_k] \in \mathbb{R}^{M \times \left(\sum_{i=1}^{k} N_{A_i}\right)}$$

where N_{A_i} is the size of the training sample subset A_i corresponding to the i-th class. So Eq. (1) is adapted to the following expression:

$$\hat{\alpha} = \arg\min_\alpha \|\alpha\|_1 \quad s.t. \quad A\alpha = y \tag{2}$$

Theoretically, the element in α represents the generalized dependence measure between the test sample y and the training sample set A, with higher numbers indicating greater similarity. Meanwhile, the recognition result α is sparse because the sparsity of CS has been inherited. Therefore, the recognition judgment depends on the residuals via each corresponding class element in

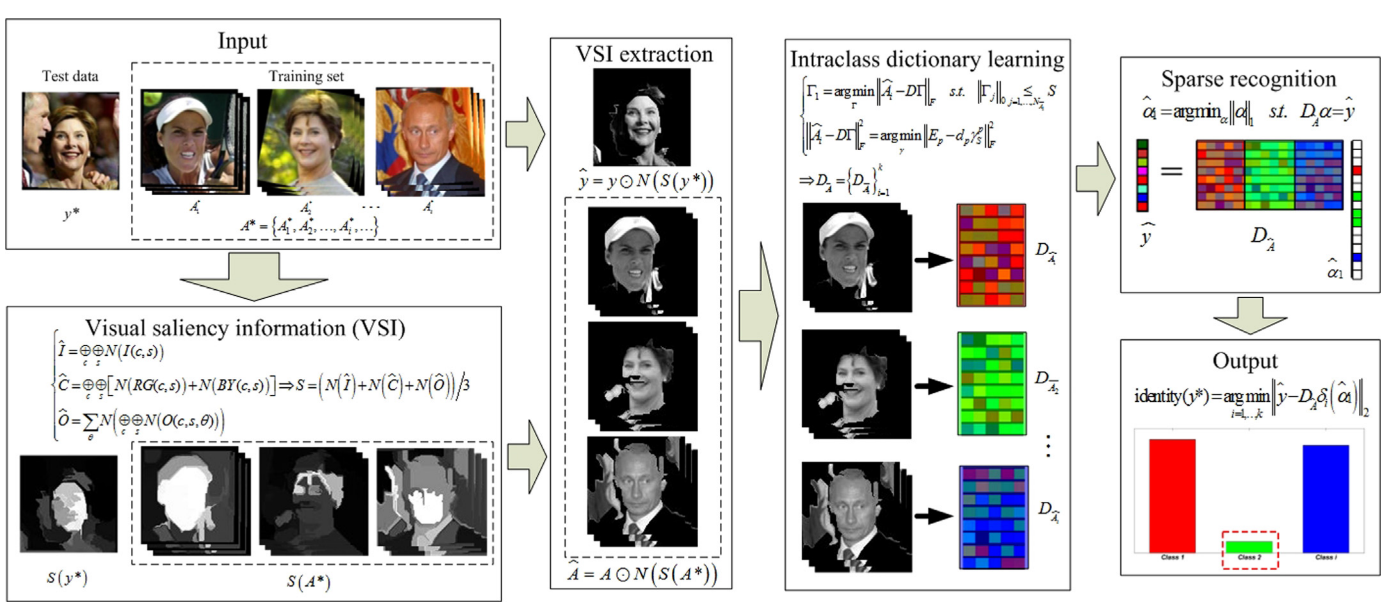

Fig. 1. Overview of our sparse recognition with saliency dictionary method.