Neurocomputing 154 (2015) 94–109

Contents lists available at ScienceDirect

Neurocomputing

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

No-reference image quality assessment with shearlet transform and deep neural networks

Yuming Li [a,*], Lai-Man Po [a], Xuyuan Xu [a], Litong Feng [a], Fang Yuan [a], Chun-Ho Cheung [b], Kwok-Wai Cheung [c]

[a] Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong Special Administrative Region, China
[b] Department of Information Systems, City University of Hong Kong, Kowloon, Hong Kong Special Administrative Region, China
[c] Department of Computer Science, Chu Hai College of Higher Education, Hong Kong, Hong Kong Special Administrative Region, China

ARTICLE INFO

Article history:
Received 29 July 2014
Received in revised form
25 September 2014
Accepted 1 December 2014
Communicated by Yongzhen Huang
Available online 15 December 2014

Keywords:
No-reference image quality assessment
Shearlet transform
Stacked autoencoders
Softmax classification

ABSTRACT

Nowadays, Deep Neural Networks have been applied to many applications (such as classification, denoising and inpainting) and achieved impressive performance. However, most of these works pay much attention to describe how to construct the relative framework but ignore to provide a clear and intuitive understanding of why their framework performs so well. In this paper, we present a general-purpose no-reference (NR) image quality assessment (IQA) framework based on deep neural network and give insight into the operation of this network. In this NR-IQA framework, simple features are extracted by a new multiscale directional transform (shearlet transform) and the sum of subband coefficient amplitudes (SSCA) is utilized as primary features to describe the behavior of natural images and distorted images. Then, stacked autoencoders are applied as 'evolution process' to 'amplify' the primary features and make them more discriminative. Finally, by translating the NR-IQA problem into classification problem, the differences of evolved features are identified by softmax classifier. Moreover, we have also incorporated some visualization techniques to analysis and visualize this NR-IQA framework. The resulting algorithm, which we name SESANIA (ShEarlet and Stacked Autoencoders based No-reference Image quality Assessment) is tested on several database (LIVE, Multiply Distorted LIVE and TID2008) individually and combined together. Experimental results demonstrate the excellent performance of SESANIA, and we also give intuitive explanations of how it works and why it works well. In addition, SESANIA is extended to estimate quality in local regions. Further experiments demonstrate the local quality estimation ability of SESANIA on images with local distortions.

1. Introduction

Visual quality measurement is a vital yet complex work in many image and video processing applications. According to the dependency of reference images, the objective image quality assessment (IQA) methods are divided into three types: full-reference (FR), reduced-reference (RR) and no-reference (NR). In FR-IQA and RR-IQA methods, the whole reference images or partial information of the reference images are assumed to be available. Since information about original image are available, state-of-the-art FR-IQA methods, such as IFC [1], VIF [2] and FSIM [3], can achieve a very high correlation with human perception. However, in many practical applications the availability of the full or partial reference image's information may be very expensive or even impossible. Because of these drawbacks, NR-IQA (or blind IQA) method has recently received a great deal of attention.

Most of the conventional NR-IQA algorithms can be classified into three types as (1) Distortion-specific, (2) Natural scene statistics (NSS), and (3) Training-based. For the first type, the distortion-specific based NR-IQA algorithms usually calibrate some specific distortions, such as JPEG [4], JPEG2000 [5]. Since this kind of NR-IQA method usually implies some prior information about distortions, it is very hard to generalize them to other new distortion types. For the second type of NSS based approaches, these NR-IQAs depend on the fact that natural scenes belong to a small set in the space of all possible signals and most distortions that are prevalent in image/video processing systems destroy the specific features of natural scenes. Resent works about this type algorithms focused on developing advanced statistical models to describe the properties of natural images, and then the blind measurement of NR-IQA is achieved by measuring the variation in terms of NSS. For example, Lu et al. [6] improved the NSS model by contourlets. Moorthy et al.

* Corresponding author.
E-mail address: yumingli4-c@my.cityu.edu.hk (Y. Li).

http://dx.doi.org/10.1016/j.neucom.2014.12.015
0925-2312/

[7] proposed BIQI, which extracted features of NSS in the wavelet domain. Saad et al. [8] proposed BLIINDS-II that applied a NSS model of discrete cosine transform coefficients. Mittal et al. [9] proposed BRISQUE that promotes extracting NSS features from the spatial domain. For the third type of training-based NR-IQA algorithms, they usually rely on a number of features extracted from images. In which a regression model is learned based on these features and labels to predict image quality. Resent works about this type algorithms focused on using advanced machine learning methods to extract effective features to represent natural images and distorted images. For example, Li et al. [10] developed a NR-IQA algorithm using a general regression neural network. Ye et al. presented a NR-IQA framework based on unsupervised feature learning framework in [11] and a NR-IQA method based on Convolutional Neural Networks in [12].

In this paper, a new NR-IQA with use of both NSS and Training-based approaches is proposed, which is named as SESANIA (ShEarlet and Stacked Autoencoders based No-reference Image quality Assessment). The proposed algorithm is a general-purpose NR-IQA, which evaluates the image quality without incorporating any prior knowledge about distortion types. Different from our previous works [13], SESANIA does not directly use the property of NSS model in shearlet domain to construct a predictor, but utilizes the sum of subband coefficient amplitudes (SSCA) as primary features to describe the behavior of natural images and distorted images. Besides, training and learning methods are also adopted through the entire framework to achieve this new NR-IQA. The main idea of SESANIA is based on the finding that the statistical property of most natural images in shearlet domain is relatively constant. Nevertheless, distorted images usually contain more or less spread discontinuities in all directions. That is, real-world image distortions disturb the natural image statistical property and discriminate it from natural images to distorted images. Shearlets are apt at precisely detecting and locating these discontinuities or singularities. Therefore, these variations in statistical property can be easily described by shearlets and applied to describe image quality distortion.

Specifically, for natural images, the SSCA in different scales has relatively constant relationship in shearlet domain. However, this constant relationship will be disturbed if a natural image is distorted by some common distortions. Motivated by this idea, SSCA can act as a primary feature descriptor to describe an image. Thus, natural images and distorted images can be distinguished by these primary features. Recently, deep neural networks have received a great deal of attention and achieved great success on various applications, such as denoising [14,15], inpainting [15], classification [16] and natural language processing [17]. In this work, we explore applying stacked autoencoders as 'evolution process' to 'amplify' the primary features and make them more discriminative. Through this evolution process, the discriminative parts of the primary features are exaggerated. Finally, by translating the NR-IQA problem into classification problem, the differences of evolved features can be easily identified by Softmax classifier. In the implementation process, SESANIA does not incorporate any prior knowledge about distortions, which makes it suitable to many distortions and easy to extend.

The remainder of the paper is organized as follows. Section 2 introduces the detailed implementation and related techniques about SESANIA. In Section 3, experimental results and a thorough analysis of this NR-IQA framework are presented. Finally, conclusion and future works are given in Section 4.

2. Methodology

The proposed framework of using deep neural network for NR-IQA is illustrated in Fig. 1. The major components in this framework include: (1) SSCA extraction in shearlet domain, (2) feature evolution using stacked autoencoders, (3) evolved feature identification using softmax classifier, and (4) quality score calculation. More details will be described in the following sub-sections.

2.1. Shearlet transform

The proposed NR-IQA is based on the shearlet transform [18–24]. This multiscale transform is a multidimensional edition of the traditional wavelet transform [25–27], and is capable for addressing anisotropic and directional information at different scales. When the dimension $n = 2$, the affine systems with composite dilations are the collections of the form:

$$SH_{\phi}f(a,s,t) = <f, \quad \phi_{a,s,t}>, \quad a>0, \quad s \in R, \quad t \in R^2 \tag{1}$$

where the analyzing factor $\phi_{a,s,t}$ is called shearlet coefficient, which is defined as

$$\phi_{a,s,t}(x) = \left|\det M_{a,s}\right|^{-\frac{1}{2}}\phi(M_{a,s}^{-1}x-t) \tag{2}$$

where $M_{a,s} = B_s A_a = \begin{pmatrix} a & \sqrt{a}s \\ 0 & \sqrt{a} \end{pmatrix}$, and $A_a = \begin{pmatrix} a & 0 \\ 0 & \sqrt{a} \end{pmatrix}$, $B_s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$. A_a is the anisotropic dilation matrix and B_s is the shear matrix. The analyzing functions associated to the shearlet transform are anisotropic and are defined at different scales, locations and orientations. Thus, shearlets have the ability to detect directional information and account for the geometry of multidimensional functions, which overcome the limitation of the wavelet transform.

Shearlets have a lot of very good mathematical properties [19]. For examples, shearlet is well localized (which means they are compactly supported in the frequency domain and have fast decay in the spatial domain), highly directional sensitivity and optimally sparse.

In summary, shearlets form a tight frame of well-localized waveforms, at various scales and directions, and are optimally sparse in representing images with edges. With these good properties, shearlets can provide more additional information about distorted images than the traditional wavelets and is suitable to NR-IQA work.

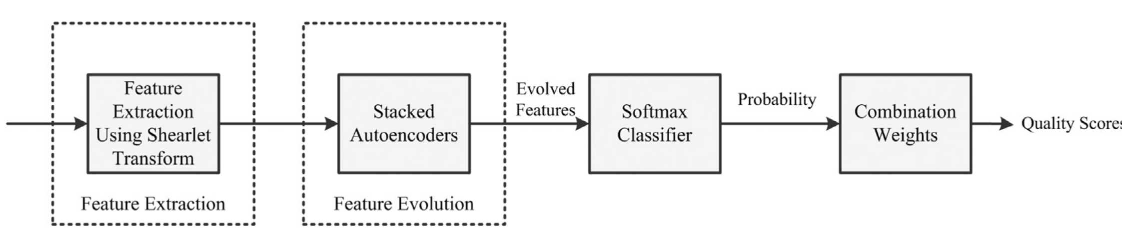

Fig. 1. Overview of the SESANIA framework.