International Journal of Data Analytics
Volume 3 • Issue 1

Convolutional Locality-Sensitive Dictionary Learning for Facial Expressions Detection

Benjamin Ghansah, University of Education, Winneba, Ghana*
https://orcid.org/0000-0002-1599-6301

ABSTRACT

Facial expression (FE) detection is a popular research area, particularly in the fields of image classification, pattern recognition, and computer vision. Sparse representation (SR) and dictionary learning (DL) have significantly enhanced the classification performance of image recognition and also resolved the problem of the nonlinear distribution of face images and its implementation with DL. However, the locality structure of face image data containing more discriminative information, which is very critical for classification, has not been fully explored by state-of-the-art SR-based approaches. Furthermore, similar coding results between test samples and neighboring training data, contained in the feature space, are not being fully realized from the image features with similar image categorizations to effectively capture the embedded discriminative information. In an attempt to resolve the foregoing issues, a novel DL method was proposed, Convolutional Locality-Sensitive Dictionary Learning (CLSDL) for facial expression detection.

KEYWORDS

Convolutional Neural Network, Dictionary Learning, Facial Expression, Feature Extraction, Locality Sensitive Adaptor, Sparse Representation

1. INTRODUCTION

Facial Expression is usually referred to as the transformation of the human face caused by an automatic response to an emotional state which may be as a result of voluntary action, in most cases spontaneous and uncontrollable. Thus, a reflex action that requires applications of a machine or intelligent system to recognize mainly the seven prototypical expressions put up by the face under any circumstance. FE has been an active research hotspot in the fields of Computer Vision (Aneja, Colburn, Faigin, Shapiro, & Mones, 2016), Pattern Recognition (Manivannan et al., 2016; Samarasinghe, 2016), Image Classification, and related fields in recent times (Kamarol, Jaward, Kälviäinen, Parkkinen, & Parthiban, 2017; Mary & Jayakumar, 2016; Valstar et al., 2017). These studies are rapidly increasing on the account of a visible manifestation of one's affective state, cognitive activity, intention, personality, and psychopathology as opined by (Martinez, Valstar, Jiang, & Pantic, 2017). FEs are composed of macro- expressions, expressed in the form of anger, disgust, fear, happiness, sadness or surprise, and other involuntary rapid facial patterns, initiating a non-rigid human face signal. Facial dynamics are computed by either observing the face signals, for example (a surprise gesture can be expressed by opening the eye wide and tightening the eyelids shows gladness) or the strain magnitude estimated using the central difference technique over the dense optical flow field detected in several regions (forehead, mouth, chin, cheek) on each subject's face (Matthew Shreve, 2011.).

DOI: 10.4018/IJDA.297520
*Corresponding Author

Lopes, A.T., et al opined that graphical images speak a thousand words, but a typical human face will speak louder since the FE is a *window* to human personality; emotions, and thoughts, thus making it play a crucial role in social communication (Lopes, de Aguiar, De Souza, & Oliveira-Santos, 2017). Note, one of the fundamental things humans learn first is, recognition of faces and interpretation of a few basic emotions from them. It is hard to imagine the expression of humor, love, appreciation, grief, enjoyment, or regret without FE. Facial recognition is one of the fastest-growing technologies, yet the least interphase with biometrics compared to other techniques such as fingerprint, iris, and gait recognition (Solanki & Pittalia, 2016). For example, in surveillance systems, instead of requiring people to place their hands on a reader (fingerprinting) or precisely positioning their eyes in front of a scanner (iris recognition), face recognition systems discreetly take pictures of individuals' faces as they enter a specified area. There is no incursion or capture delay since the subjects are entirely oblivious of the actions, hence they (subjects) do not feel monitored under surveillance or their privacy invaded.

Expression recognition is a task that humans perform daily and effortlessly but that is not yet easily performed by computers without any challenge. These challenges include but are not limited to the different poses of facial images, partial occlusion, illumination variations, and facial expressions (e.g. facial features from one subject in two different expressions as expressed in fig. 1 below) which is why FE recognition remains a daunting task. In addition, certain expressions like ”sad” and ”fear” are in some cases very similar (Martinez & Valstar, 2016), which Darwin suggested manifests in animals as well (Darwin & Prodger, 1998). Hence, efficiently recognizing the different FEs is important for both the evaluation of emotional states and the automated face recognitions.

Figure 1 has three subjects who showed signs of *happy expression*. It could be observed that images differ a lot from each other not only in the way that the subjects show their expression but also in the lighting, background, pose, and brightness that is associated with the image. The major aim of the paper is to detect varying FEs and variability in face appearance which could be caused by changes in facial expressions which could be as a result of stimulation by varying individuals' emotional states. Here we propose to incorporate CNN into the structure of the Locality Sensitive Dictionary Learning method.

Figure 1. Similar Facial Expression being expressed by three different subjects. The images are from JAFFE and CK+ databases (Naidoo, Tapamo, & Khutlang, 2018).

FE recognition systems are mainly split into two main categories: static-based method in (Ali, Iqbal, & Choi, 2016; M. Liu, Li, Shan, & Chen, 2015; P. Liu, Han, Meng, & Tong, 2014; Song, Kim, & Jeon, 2014; S. Zhang, Zhao, Chuang, Guo, & Chen, 2017) and dynamics-based method discussed in (Byeon & Kwak, 2014; Fan & Tjahjadi, 2015; W. Zhang, Zhang, Ma, Guan, & Gong, 2015). Static-based methods use permanent information (feature vectors) of data from present image inputs