ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICRITCSA

M S Ramaiah Institute of Technology, Bangalore

Vol. 5, Special Issue 2, October 2016

Facial Image Analysis: Recent Trends and Approaches

Soumya C S[1], Dr Thippeswamy G[2]

M.Tech Student, Dept. of CSE, BMSIT&M, Bangalore, India[1]

Prof and HOD, Dept. of CSE, BMSIT&M, Bangalore, India[2]

Abstract: In recent years face recognition has received substantial attention from both research com-munities and the market, but still remained very challenging in real applications. A lot of face recognition algorithms, along with their modifications, have been developed during the past decades. Deep learning has recently achieved very promising results in a wide range of areas such as computer vision, speech recognition and natural language processing. It aims to learn hierarchical representations of data by using deep architecture models. Facial emotion recognition is one of the most important cognitive functions that our brain performs quite efficiently. State of the art facial emotion recognition techniques are mostly performance driven and do not consider the cognitive relevance of the model. Similarly, Facial image analysis through 3D spectral information is gaining lot of scope. Hyperspectral cameras provide useful discriminants for human face recognition that cannot be obtained by other imaging methods. Hence, the facial analysis through Hyperspectral imaging is a great advantage. In this paper, we try to comprehend the recent emerging technologies in the field of image analysis for faces.

Keywords: Face recognition, image analysis, 3D image, Deep Learning, Hyperspectral Imaging.

I. INTRODUCTION

In recent years face recognition has received substantial attention from researchers in biometrics, pattern recognition, and computer vision communities. The machine learning and computer graphics communities are also increasingly involved in face recognition. This common interest among researchers working in diverse fields is motivated by our remarkable ability to recognize people and the fact that human activity is a primary concern both in everyday life and in cyberspace.

Besides, there are a large number of commercial, securities, and forensic applications requiring the use of face recognition technologies. These applications include automated crowd surveillance, access control, mug shot identification (e.g., for issuing driver licenses), face reconstruction, design of human computer interface (HCI), multimedia communication (e.g., generation of synthetic faces), and content-based image database management. A number of commercial face recognition systems have been deployed, such as Cognitec, Eyematic, Viisage, and Identix. Facial scan is an effective biometric attribute/indicator. Different biometric indicators are suited for different kinds of identification applications due to their variations in intrusiveness, accuracy, cost, and ease of sensing.

II. SURVEY

To attack a pose-invariant face recognition task, one possible remedy approach is to capture multi-view face images from each individual and estimate all the other possible pose positions. However, it is often not practical to collect multi-view images for each individual in real applications. As a result, the virtual view synthesis scenarios, which base on 2D pose transformation or 3D face reconstruction, are proposed to substitute the demand of real views from limited known views (i.e. only the frontal view in our framework) [1].

Deep Learning methods have performed very well in digit recognition dataset [2]. Our setting is very similar to the task of digit recognition. Corresponding to the digit labels we have emotion labels. But emotion recognition is much more complicated because digit images are much simpler than face images depicting various expressions. Moreover the variability in the images due to different identities hampers the performance. Human accuracy in facial expression recognition is not as good as in digit recognition and is also aided by other modes of information such as context, prior experience, speech among others.

Spectroscopy is a valuable tool for a large number of applications [3]. Spectral measurements from human tissue, for example, have been used for many years for characterization and monitoring applications in biomedicine. In remote sensing, researchers have shown that Hyperspectral data are effective for material identification in scenes where other sensing modalities are ineffective. The introduction of Hyperspectral cameras has

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICRITCSA

M S Ramaiah Institute of Technology, Bangalore

Vol. 5, Special Issue 2, October 2016

led to the development of techniques that combine spectral and spatial information.

As Hyperspectral cameras have become accessible, computational methods developed initially for remote sensing problems have been transferred to biomedical applications. Considering the vast person-to-person spectral variability for different tissue types, Hyperspectral imaging has the ability to improve the capability of automated systems for human identification [4].

Current face recognition systems primarily use spatial discriminants that are based on geometric facial features. Many of these systems have performed well on databases acquired under controlled conditions.

III. 3D SPECTRAL IMAGING

3D modeling is an essential part of many applications such as Virtual Reality VR, Virtual Museum, Urban scenes, 3D games, and many other applications. However in order to increase the realism of any model a texture maps have been used to achieve this realism.
There are many ways to produce 3D model such as 3D modeling using a graphics designer to produce such model, 3D reconstructions techniques and 3D scanning. The later method "3D scanning " is by far the most sophisticated method for creating realistic 3D models by using 3D Scanners "Digitizers" to obtain the actual model.
The use of texture maps on 3D models has become an essential process of any 3D modeling System. Texture mapping is usually done by a graphic designer in non-critical system such as Virtual Reality "VR" and video games. But for critical systems where the texture mapping should be 100% accurate such as medical applications and in large projects the need for automation of the process becomes a must.
However the process of automating the texture mapping in robust way is yet to be achieved. The main problem in texture mapping is in knowing which pixel on the texture image corresponds to which vertex on the 3D model. Many methods and algorithms have been proposed to achieve the automation of texture mapping.

3-D face recognition is defined as a biometric technique which uses individual 3-D facial shape to recognize human faces using 3-D models of both probe and gallery faces. A 3-D human face captured during face acquisition may contain unwanted body parts or areas like hair, ears, neck, shoulders and accessories like glasses and ornaments that need to be effectively eliminated. Major landmarks facilitate the segmentation process which extracts face shape from the entire scan. Facial shape needs to be aligned before actual matching.
3-D registration techniques are used for face shape alignment. The discriminant features are extracted and stored using the surveyed region-based and holistic approaches for all faces in the gallery. Feature vector for a probe (query) scan is extracted and matched against the gallery feature vectors one by one. The gallery scan that has the closest matching distance with the probe scan below a predefined distance threshold is considered as the recognized gallery scan, and the class of the corresponding feature vector is declared as the result.

IV. RECENT TRENDS

4.1 3d Facial Image Analysis
We now review the prior work in generic face alignment, pose-invariant face alignment, and 3D face alignment. The first type of face alignment approach is based on Constrained Local Model (CLM), where an early example is ASM. The basic idea is to learn a set of local appearance models, one for each landmark, and the decisions from the local models are fused with a global shape model. There are generative or discriminative approaches in learning the local model, and various approaches in utilizing the shape constraint. While the local models are favored for higher estimation precision, it also creates difficulty for alignment on low-resolution images due to limited local appearance. In contrast, the AAM method and its extension learn a global appearance model, whose similarity to the input image drives the landmark estimation [5]. While AAM is known to have difficulty with unseen subjects, the recent development has substantially improved its generalization capability. Motivated by the Shape Regression Machine in the medical domain, cascaded regressor-based methods have been very popular in recent years. On one hand, the series of regressors progressively reduce the alignment error and lead to a higher accuracy. On the other hand, advanced feature learning also renders ultra-efficient alignment procedures. Other than the three major types of algorithms, there are also works based on deep learning, graph-model, and semi-supervised learning.

Despite the explosion of methodology and efforts on face alignment, the literature on pose-invariant face alignment is rather limited. There are four approaches explicitly handling faces with a wide range of poses. Zhu and Ranaman propose the TSPM approach for simultaneous face detection, pose estimation and face alignment [6] . An AFW dataset of in-the-wild faces with all poses is labeled with 6 landmarks and used for experiments. The cascaded deformable shape model (CDM) is a regression-based approach and probably the first approach claiming to be "pose-free", therefore it is the most relevant work to us. However, most of the experimental datasets contain near-frontal view faces, except the AFW dataset with improved performance than. Also, there is no visibility estimation of the 2D landmarks. Zhang et al. develop an effective deep learning based method to estimate 5 landmarks. While accurate results are obtained, all testing images appear to

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICRITCSA

M S Ramaiah Institute of Technology, Bangalore

Vol. 5, Special Issue 2, October 2016

be within ~+-60 so that all 5 landmarks are visible and there is no visibility estimation. The OSRD approach has the similar experimental constraint in that all images are within +-40^0. Other than these four works, the work on occlusion-invariant face alignment are also relevant since non-frontal faces can be considered as one type of occlusions, such as RCPR nd CoR. Despite being able to estimate visibilities, neither method has been evaluated on faces with large pose variations. Finally, all aforementioned methods in this paragraph do not explicitly estimate the 3D locations of landmarks.

3D face alignment aims to recover the 3D locations of facial landmarks given a 2D image. There is also a very recent work on 3D face alignment from videos. However, almost all methods take near-frontal view face images as input, while our method can handle faces at all poses. A relevant but different problem is 3D face reconstruction, which recovers the detailed 3D surface model from one image, multiple images, or an image collection. Finally, 3D face model has been used in assisting 2D face alignment. However, it has not been explicitly integrated into the powerful cascaded regressor framework, which is one of the main technical novelties of our approach.

4.2 Deep Learning

State of the art approaches in facial emotion recognition use Active Appearance Models (AAMs), FACS labels or some other sophisticated feature extraction scheme. AAMs can be learned from a set of training images and can be fitted on a new face to generate the landmark positions which can further be used to design features [7]. Thus, in an automatic setting either the availability of landmark point on face images is assumed or can be obtained by fitting the model. FACS labels attempt to decompose human emotions in terms of Action Units (AUs) which correspond to specific muscle movements. FACS coding system is used in psychology and animation to classify facial expressions in a consistent and systematic manner. But as of now FACS labels can only be given by experts or trained individuals.

One problem with ad hoc feature extraction schemes is that we need to design separate feature extraction mechanism for each visual task to be performed. Therefore, it makes much more sense to have generic scheme for learning what transformations in the input space may lead to good features for performing a particular task [8].

There is ample evidence that our visual processing architecture is organized in different levels. Each level transforms the input in a manner that facilitates the visual task to be performed. Another appealing feature of deep learning models is that there can be feature or sub-feature sharing. Computationally also, it has been shown that insufficiently deep architectures can be exponentially inefficient.

Deep Learning methods have performed very well in digit recognition dataset. Our setting is very similar to the task of digit recognition. Corresponding to the digit labels we have emotion labels. But emotion recognition is much more complicated because digit images are much simpler than face images depicting various expressions. Moreover the variability in the images due to different identities hampers the performance. Human accuracy in facial expression recognition is not as good as in digit recognition and is also aided by other modes of information such as context, prior experience, speech among others.

Restricted Boltzmann Machine (RBM) is only interesting because they can be efficiently stacked up layer by layer to form a deep network. First an RBM is trained on the visible layer. Once trained the weights are frozen and the hidden layer activations act as the input for the next RBM. Thus a Deep Belief Networks (DBN) with any number of layers can be formed by stacking RBMs as mentioned above. It has also been shown that increasing the number improves the variation lower bound on the probability of the training data. RBM acts as a fundamental unit in the whole DBN. There are other models that can be used instead of RBMs such auto encoders (sparse), denoising auto encoders. Details about sparse auto encoder can be found. Once the layer by layer model has been trained a final supervised fine tuning step which adjusts the weights to improve the performance on the particular task in hand [9].

4.3 Hyperspectral Imaging

Hyperspectral cameras provide useful discriminants for human face recognition that cannot be obtained by other imaging methods. The Hyperspectral images were collected using a CCD camera equipped with a liquid crystal tunable filter to provide 31 bands over the near-infrared (0.7 /spl mu/m-1.0 /spl mu/m). Spectral measurements over the near-infrared allow the sensing of subsurface tissue structure which is significantly different from person to person, but relatively stable over time. The local spectral properties of human tissue are nearly invariant to face orientation and expression which allows Hyperspectral discriminants to be used for recognition over a large range of poses and expressions [10].

Several of the limitations of current face recognition systems can be overcome by using spectral information. The interaction of light with human tissue has been studied extensively by various researchers and determines tissue spectral properties. The epidermal and dermal layers of human skin constitute a scattering medium that contains several pigments such as melanin, hemoglobin, bilirubin, and carotene.

Small changes in the distribution of these pigments induce significant changes in the skin’s spectral reflectance. The effects are large enough, for example, to enable algorithms for the automated separation of melanin and hemoglobin

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICRITCSA

M S Ramaiah Institute of Technology, Bangalore

Vol. 5, Special Issue 2, October 2016

from RGB images. Recent research has measured skin reflectance spectra over the visible wavelengths and proposed models for the spectra. Other researchers have used a skin reflectance model over the 0.3 m-0.8m range to propose a method for skin [11].

V. CONCLUSION

This paper presents a method to enhance the perceived quality of a face in an image using physical parameters instead of directly modifying the image data. Thus we can conclude that the recent trends and approaches in the facial image analysis are the 3D spectral imaging, Deep Learning and Hyperspectral imaging.

REFERENCES

[1] Maxime Devanne, Hazem Wannous, Stefano Berretti, "3-D Human Action Recognition by Shape Analysis of Motion Trajectories on Riemannian Manifold", pp-2168-2267, 2014, IEEE.

[2] B H Shekar, G.Thippeswamy, P Shivakumara, "A Rule based Model for Efficient Representation and Accurate Recognition of Human Faces", International Conference on Advances in Computer Engineering,2010.

[3] Seyedehsamaneh Shojaeilangari, Karthik Nandakumar "Robust Representation and Recognition of Facial Emotions Using Extreme Sparse Learning", pp- 1057-7149 (c) 2015, IEEE.

[4] Showing love, Feipeng DA, Xing Deng, "A 3D Face Recognition Method Using Region-Based Extended Local Binary Pattern", 978-1-4799-8339-1/15/, ICIP 2015.

[5] Manar D. Samad and Khan, M. Iftekharuddin, "Frenet Frame-Based Generalized Space Curve Representation for Pose-Invariant Classification and Recognition of 3-D Face", pp-2168-2291, 2016 IEEE.

[6] Zhongjun Wu, Jiayu Li, Jiani Hu, and Weihong Deng, "Pose-Invariant Face Recognition Using 3D Multi-Depth Generic Elastic Models", pp-4789-6026-15, 2015, IEEE.

[7] Y. Lei, M. Bennamoun, M. Hayat, and Y. Guo, "An efficient 3D face recognition approach using local geometrical signatures," Pattern Recog., vol. 47, no. 2, pp. 509–524, Feb. 2014.

[8] R. Slama, H. Wannous, and M. Daoudi, "3D human motion analysis framework for shape similarity and retrieval," Image Vis. Comput., vol. 32, no. 2, pp. 131–154, 2014.

[9] H. Drira, B. B. Amor, A. Srivastava, M. Daoudi, and R. Slama, "3D face recognition under expressions, occlusions, and pose variations," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 9, pp. 2270–2283, Sep. 2013.

[10] Georey E. Hinton, Yee-Whye Teh and Simon Osindero, A Fast Learning Algorithm for Deep Belief Nets. Neural Computation, pages 1527-1554, Volume 18, 2008.

[11] Susskind, J.M. and Hinton, G.E. and Movellan, J.R. and Anderson, A.K., Generating facial expressions with deep belief nets, Affective Computing, Emotion Modelling, Synthesis and Recognition, pages 421-440,2009.