REGION OF INTEREST FEATURE EXTRACTION IN FACIAL EXPRESSIONS WITH CONVOLUTIONAL NEURAL NETWORK CLASSIFICATION

B. Rajesh Kumar
Department of Computer Science and Engineering, Karpagam Academy of Higher Education, India

Abstract

This article introduces a face expression using a processing sequence that involves the CNN 3D Skull Description ROI point extraction. The first step is the noise reduction process using Gaussian filtering from the input 3D face image. The functionality is extracted with the Otsu Thresholding Multilevel Operation. Using morphological erosion, boring off, and region operators, the derived features from the photos are improved. The extraction of points is then sent to the Interest Area (ROI) which essentially extracts points in both photos. The experiment is performed on a sequence of 100 photographs, including the cranial and facial pictures. The extracted ROI of the enriched input image is sent to the CNN, which is training and testing new input images directly applied to the CNN classification. The trained ROIs are transmitted as trained data. For 3D modelling, the CNN classification is used and finally, face expression is used to compare the similitude between the original image and the image being tested. The result shows that a higher classification rate for the proposed method is achieved, efficiently extracting the points and improving the classification required for 3D modelling than existing methods.

Keywords:

Facial expression, CNN Classification, ROI Extraction, 3D Images

1. INTRODUCTION

Facial expression is a method by which photos of a skull are contrasted with facial images of a once-living individual who could represent the skull [3]. This method is common when the samples of a certain missing person are suspected to be linked to, for whom photos are visible. The methodology and the nature of the applications have evolved significantly in the last two decades in particular. Specialists in forensics and forensics, data and morphometric analyses were the driving forces behind these developments. [1].

The attempt made by inspecting the skull removed to approximate a human visual appearance is facial rebuilding [8] – [13]. It is used, in the forensic sense, when a crane has been recuperated or attempts made to relate to traditional methods which are regarded as new and thus of forensic significance. Facial restoration is then used to produce a face-lift and present information to the general public about missing persons, primarily in the newspapers [6]. Farther, facial expression is one of the long-lasting tasks for the researchers. But usually the expert uses a different procedure [7]. There is no systematic solution. Therefore, it is of great importance to design automated processes that enable the forensic anthropologist to implement them [2].

This paper presents a face formulation that includes CNN 3D modelling classification, ROI point extraction. The first step is the noise reduction process using Gaussian filtering from the input 3D face image [4]. The functionality is extracted with the Otsu Thresholding Multilevel Operation. Using morphological erosion, boring off, and region operators, the derived features from the photos are improved [5]. It is then sent to the ROI point extraction, where the points in both images are efficiently extracted. The experiment is performed on a sequence of 100 photographs, including the cranial and facial pictures. The extracted ROIs from the improved input image are sent as a trained data to the CNN, which is trained and tested directly on the CNN-classifier input image. For 3D rendering, the CNN classification is used and eventually, face expression is used to compare the similitude between the original image and the image being studied.

2. PROPOSED METHOD

The Fig.1 gives the architecture of the proposed method and sub-sections are used to give the complete processing.

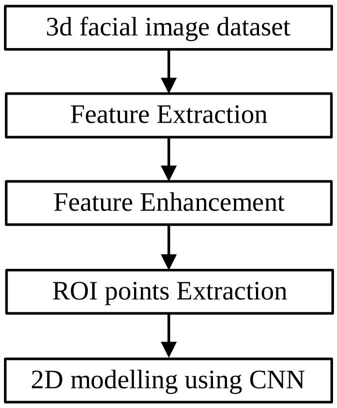

Fig.1. Architecture of the Proposed Method

2.1. PRE-PROCESSING

Filtered with Gaussian filters [12], the input of 3D facial and 3D cranial image is used to eliminate noise from the CT image.

ROI Extraction: This research uses the extraction of points or point seed points from the 3D facial and 3D cranial pictures from Area of Interest (ROI). The attribute points are clearly distinguished for the segmentation of both 3D facial and 3D cranial images. The position of the ROI points are estimated when the target area is extracted, provided that both 3D and 3D cranial images contain a ROI point. The two images initially identify the target region and ultimately extract the characteristic ROI points. At the beginning, both techniques are used in cranial photographs, so they represent the face topology of the captured image that can be sent as the reference image. The operation is then done in the image of the face. Therefore, instead of bringing the whole pixel processing, the analysis extracted the ROI points from both 3D images. The processing of the 3D binary cranial image and the like process for the 3D binary image is taken into consideration in the following process.

At the beginning, the enhanced image feature is considered to be the ROI extraction input that is purely binary. The pixels within the cranial area are treated as "1" and vice versa. The treatment of the pixels is iterated to the complete pixel.

The Harris operator is used to generate the ROI points during the last step of ROI extraction. The 3D cranial image corners of the Harris are stable, even though the form of the target area is different. The ROI obtained is the target area of the 3D face image obtained.

2.2. 3D MODELLING

The points are located on the two pictures after the ROI points are separated from the 3D facial and 3D cranial. The differences between the ROI of 3D facial and 3D cranial are compared after the points are displayed and the skull is modelled using 3D face and 3D cranial images extracted from the points. The ROI of multiple extracted images of both a 3D-facial and a 3D-cranial image set is given as a CNN input for the training of this profound study algorithm with an input of about 100 images.

Facial Expression: Facial expression is the process by which images from the CNN algorithm are compared to the original image. If the difference between the ROI of 3D face and 3D cranial is large, the ROI points of the face are extracted once more by redirecting the ROI extraction process across that particular area, and if the difference between them is less, they are used for two-dimensional modelling of the cranial.

2.3. CNN CLASSIFICATION

In this paragraph we present mainly CNN building blocks, i.e. the overall and pooling layers that represent CNNs, the simple objects coded by GA. In specific, filters are used to conduct convolutional operations on the input data on the convolutional sheet. The matrix can be seen as one filter. The filter slides horizontally (with a certain step size) during convolutional processing, then moves vertically (with another step size) for the next horizontal slide before the full picture has been scanned. A new matrix named the function map forms the set of filter outputs. The horizontal and vertical step sizes are referred to as step width and height. A parameter in a corresponding CNN architecture is the exact number of function maps used. Moreover, there are two convolutional operations: the same convolutional operation padding zeros to the input data if the region to overlap for the filter is not appropriate, and the valid convolutional operation that pads nothing. The number of function maps, the size of the filter, the phase size and the convolutional working kind are thus the parameters of an overall layers. There are several typical components in a pooling layer with the except that (1) the kernel is called the kernel that does not have a value, (2) the kernel output is the maximum or average value of the region where it is delayed, and (3) the input data is not affected by a pooling level in space. Once the highest value is returned, the pooling layer is max, otherwise of average. Thus, the kernel size, phase size and type of pooling are all parameters of a pooling layer. Furthermore, the entirely linked layers are normally combined into a CNN.

Next we will clarify the reasons why the fully-connected layers have been discarded using two convolution layers in a skip layer and the algorithm setting of the skip and pooling layers. In specific, the tail of a CNN normally has many completely linked layers attached. However, because of its thick connection, the completely bound layer quickly induces the overfitting phenomenon. The spontaneous elimination of part of the links is widely used as a way of minimising overfitting. However, a single parameter is applied to each decrease. The promising output of the respective CNN will only result in a correctly defined parameter. In the meantime, two parameters are often difficult to tune in the number of fully connected layers and the number of neurons per each fully connected layer. The search field would greatly broaden and make discovering the right CNN architecture more challenging as fully-connected layers are embedded in the proposed coding strategy. Two convolutional layers in a skip layer have been influenced by ResNet, the usefulness of these skipping layers has been shown experimentally. But the sizes of the signature maps of each ResNet skip layer are the same. The sizes of function maps will be standardised, making them more flexible, in the proposed coding strategy. In addition, the uncertainty environment and the step would make the input data dimension the same, which with this automated design is more versatile step does not alter the image size. As the buffer, kernel and strip size settings in the pooling levels are calculated, they are all based on the principles of the current manufactured CNNs. In addition, our experience in manually tuning the architectures of the CNNs is another significant factor to define these settings.

The explanations for the creation of an asynchronous portion and cache are also given. Since training on CNNs is very time intensive and depends on the individual architecture, from many hours to even several months, it is intended to speed up fitness tests within the proposed algorithm. In other words, the asynchronous portion is, in particular, a parallel GPU-based calculation platform. Given the machine design of gradient calculations, GPUs are usually supported for deep learning algorithms to speed up training. In reality, existing libraries such as Tensorflow and PyTorch support multiple-GPU calculation. The parallel calculations are, however, focused on the parallel data and parallel model pipelines. The data input data is divided into many seven groups in the data parallel pipeline, and for the estimation, every group is positioned on one GPU. The explanation is that one GPU's restricted memory cannot accommodate all of the data simultaneously. A module is split into several small models in the model-parallel pipeline and each GPU has one small model. The explanation is that a single GPU cannot run an entire model with the minimal computational capability. The parallel pipeline designed, however, simply does not collapse into any of the pixels, but is focused on them on a higher level. Therefore, a GPU machine resource, particularly for communities, is completely exploited by this asynchronous portion. Moreover, a broad problem is always solved using the asynchronous part if the problem can be separated into several different sub-problems. The overall running time of the entire problem is thereby minimised by parallel performance of these sub-problems on various computing platforms. Evolutionary algorithms in the past have traditionally been used to address issues that do not take time for fitness assessment1 and such asynchronous components do not need a high degree of development. On occasion, the combined modules are only used depending on the languages of programming adopted. Never mind, almost every part of this type is CPU-based, and cannot train deep neural networks effectively, mostly because the acceleration platform is based on GPUs for neural networks. In addition, the individual fitness testing is autonomous and follows the scene of this process. The asynchronous component is built in the proposed algorithm, inspired by the reasons described above. This cache portion is often used for the acceleration of the fitness assessment on the basis of the following considerations: (1) persons who survive in the next generation should not reassess the fitness if the architecture is not changed, and (2) the design which was assessed could be regenerated in a further generation by the mutation and crossovers. Notice that for the second consideration the weight inherited from Large-Scale Evolution did not work.

3. RESULTS AND DISCUSSIONS

The research comprises of the 10 representations of the cranial and facial images with feedback from 100 training sets. The results are assessed for runtime and the precise classification between the methods proposed and existing approaches.

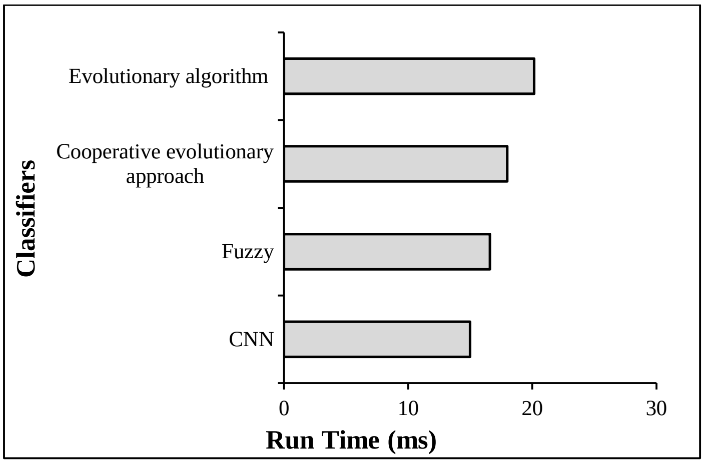

Fig.2. Run Time

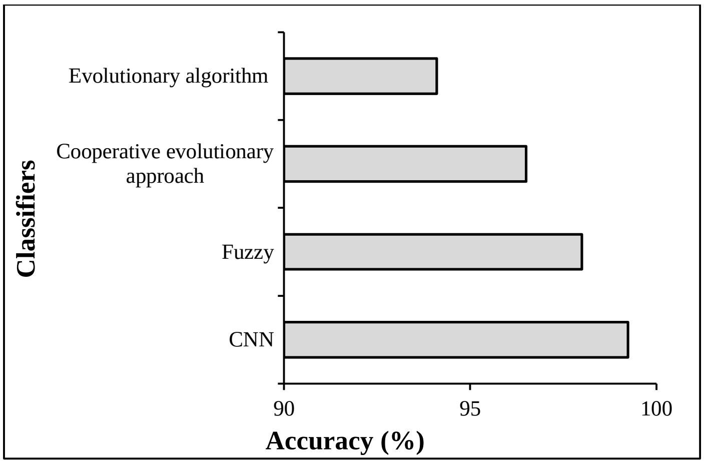

Fig.3. Classification Accuracy

This results in improving the accuracy of classification with reduced range than the current methods by the proposed method. In comparison, from the first generation to the second generation, classification accuracy is slowly growing since the population is randomly initialised at the start of growth. The progress in classification precision is much lower than the previous generations from the second to fourth generations and rises dramatically immediately afterwards to the 15th. And then the precision of the classification doesn't change before the production stops, which means that in this particular case the setting of 20 generations is appropriate since this is an environment for the proposed algorithm.

4. CONCLUSIONS

A series of treatment for the facial expression of the 3D input image is used in this paper. The noise is removed first using Gaussian filtering and then the thresholding at multilevel Otsu extracts the characteristics. The derived characteristics are improved by morphological degradation, edge-off and the opening surface operator. For classification and construction of the 3D modelling, the ROI points are then removed. For the classification of the test images, the CNN classification scheme indicates that the solution proposed is more graded than the current approaches and is easily used to model the skull. Two components in CNN were designed to accelerate the health assessment, saving a lot of computing energy. However, in the resolution of typical problems, the computing capabilities used are still fairly high than those of CNN. Several algorithms based on evolutionary calculation techniques have been developed in the area of costly optimization problems. We will in future aim to build efficacious process estimation procedures to accelerate the health assessment of CNNs considerably.

REFERENCES

[1] M. Heikkilla, M. Pietikainen and C. Schmid, "Description of interest regions with Local Binary Pattern", *Pattern Recognition*, Vol. 42, No. 3, pp. 425-436, 2009.

[2] Cong Geng and Xudong Jiang, "Fully Automatic Face Recognition Framework Based on Local and Global Features", *Machine Vision and Applications*, Vol. 24, No. 3, pp. 537-549, 2013.

[3] Hamidreza Rashidy Kanan and Karim Faez, "Recognizing Faces using Adaptively Weighted Sub-Gabor Array from a Single Sample Image Per Enrolled Subject", *Image and Vision Computing*, Vol. 28, No. 3, pp. 438-448, 2010.

[4] Yousra Ben Jemaa and Sana Khanfir, "Automatic Local Gabor Features Extraction for Face Recognition", *International Journal of Computer Science and Information Security*, Vol. 3, No. 1, pp. 1-14, 2009.

[5] G. Zhao and M. Pietikainen, "Dynamic Texture Recognition using Local Binary Patterns with an Applications to Facial Expressions", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, No. 6, pp. 915-928, 2007.

[6] Zohair Al Ameen, Ghazali Sulong, Amjad Rehman, Abdullah Al-Dhelaan, Tanzila Saba and Mznah AlRodhaan, "An Innovative Technique for Contrast Enhancement of Computed Tomography Images using Normalized Gamma Corrected Contrast-Limited Adaptive Histogram Equalization", *EURASIP Journal on Advances in Signal Processing*, Vol. 2015, No. 1, pp. 1-12, 2015.

[7] Matthew Turk and Alex P. Pentland, "Eigen Faces for Recognition", *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, pp. 71-86, 1991.

[8] Rajesh Garg, Bhawna Mittal and Sheetal Garg, "Histogram Equalization Techniques for Image Enhancement",

International Journal of Electronics and Communication Technology, Vol. 2, No. 1, pp. 107-111, 2011.

[9] Rafael C. Gonzalez and Richard E. Woods, "*Digital Image Processing*", Addison-Wesley, 1993.

[10] Thamizharasi and J.S. Jayasudha, "A Literature Survey on various Illumination Normalization Techniques for Face Recognition with Fuzzy K Nearest Neighbour Classifier", *ICTACT Journal on Image and Video Processing*, Vol. 5, No.4, pp. 1044-1051, 2015.

[11] Erik Hjelmas and Boon Kee Low, "Face Detection: A Survey", *Computer Vision and Image Understanding*, Vol. 83, No. 3, pp. 236-274, 2001.

[12] S. Ranganatha and Y.P. Gowramma, "Development of Robust Multiple Face Tracking Algorithm and Novel Performance Evaluation Metrics for Different Background Video Sequences", *International Journal of Intelligent Systems and Applications*, Vol. 10, No. 8, pp. 19-35, 2018.

[13] S. Ranganatha and Y.P. Gowramma, "Image Training, Corner and FAST Features Based Algorithm for Face Tracking in Low Resolution Different Background Challenging Video Sequences", *International Journal of Image, Graphics and Signal Processing*, Vol. 10, No. 8, pp. 39-53, 2018.