Foundation University Journal of Engineering & Applied Sciences
Vol. 1, Issue. 1. June 2020

Research article

Automatic Facial Expression Analysis and Recognition using Zone Based Active and Salient Facial Patches

Ukasha[1], Khalid Iqbal[1], Taimur Shahzad*[1]

Muhammad Faseeh[1]

[1] Department of computer science, COMSATS University Islamabad Attock campus, Pakistan

*Corresponding Author: Taimur Shahzad. Email: fa20-rcs-011@cuiatk.edu.pk

Abstract:

Recognition of facial expression has many useful applications that have drawn researcher's interest over the past decade. Extraction of features is a major step in the analysis of expression which leads to fast and accurate recognition of expression. Machine learning systems struggle to recognise facial expressions since everyone expresses themselves differently, and images of the same individual can differ in brightness, background, and location for the same expression. As a result, facial expression recognition remains a difficult computer vision task. By merging multiple techniques from computer vision and pattern recognition, we hope to develop a robust technique for automatic facial emotion analysis and recognition that uses zone-based active and salient areas of the human face. Face recognition is closely related to expression recognition, which has gotten a lot of research and a variety of algorithms. Many strategies are available for facial expression recognition (FER), which may be thought of as a subset of pattern recognition. Preprocessing, active and salient patch extraction, and classification were the four modules that we divided into when constructing a FER system. Face detection is done using the Voila Jones method, and then through facial patches, features will be retrieved. The active facial patches were found on the areas of the face that alter dramatically during various expressions. After face landmarks are detected, active patches are discovered, and hybrid characteristics are determined from these patches. The computational cost of extracting features is reduced by using small sections of the face rather than the entire face. The use of zoning has yielded impressive outcomes. Using linear discriminant analysis, the function's dimensionality is reduced, and the support vector machine is used to further define the function (SVM). Based on classification expression is recognized. We evaluated our algorithm on Extended Cohn-Kanade (CK+) dataset.

Keywords: FCR, CK, FER

1. Introduction

Expressions or emotions reflect the emotional state of a person. Facial expressions and body language tells us more than the words about one's state of mind. Facial Expression Recognition has been close to researchers for past few years due to its role in various fields like Virtual Reality, Healthcare, and Intelligent Tutoring System [1]. The goal of the Face Expression Recognition system is to use facial images to determine emotional states such as anger, fear, happiness, sadness, and surprise. In our daily lives, facial expressions are quite significant. According to study, linguistic language (verbal parts) accounts for 7% of human communication, whereas paralanguage (vocal parts) accounts for 38%, and facial expressions account for 55% [2]. Therefore, facial expressions play a vital role in perceiving information in face-to-face communication.

DOI: *10.32604/x.2021*

Significant progress has been achieved in this field in recent years, with better face models and more powerful processors allowing face recognition systems to perform well in limited scenarios [3]. Face recognition is still difficult in general situations since face images are influenced by a variety of elements such as brightness, head attitude, expression, and so on. From the perspective of computer vision [4], the most difficult of all these noises is facial expression, because expressions genuinely change the three-dimensional object. Other elements, like as brightness and position, only have an impact on image parameters. To remove "expression noise," one must first estimate an image's expression, which is referred to as "Facial Expression Recognition." [5].

It is a difficult task to automatically recognize facial expressions with accuracy. Firstly, it's hard to find the similarity of same emotions between different persons because they may express the same emotions in different way and secondly some expressions are difficult to differentiate so, it's also hard to find the difference between expressions of same person.[6]. Our expressions are made by the facial muscle's activation. These expressions or emotions are complex in nature often consists of information about one's state of mind. Automatic recognition of facial expressions can be a useful component of Human-Computer interaction [7]. It can also be used in the field of behavioral science analysis or in clinical practices. It has been the researcher's interest for long period of time and have progress in recent decades [8]. Though a large progress has been made but it remains a challenging and difficult task to recognize facial expressions with high accuracy because of facial expressions complexity [9]. Facial Expression Recognition system basically contains four basic modules which are Preprocessing, Feature Extraction, Feature Selection, and Recognition.

Because the face carries most of the expression-related information, the face is recognised and cropped from the given input image in the preprocessing module, and the image is then run through the Gaussian Filter [9] for details improvement. We will be using Voila Jones method for the detection of the face. It is the most common method used for this purpose. In feature extraction module we extract features of each expression in the given images. There are many methods used for this purpose like Local Binary Patterns, Independent Component Analysis (ICA), Local Feature Analysis and Discrete Cosine Transform (DCT) etc. Our next step is to select features from the extracted features because we only need features that are important in transformation of one expression into another. After that in Recognition step classifier is firstly trained with training data and then it is used to generate the labels for facial expressions contained in the input image [10].

In this research, we propose a facial landmark detection technique and a facial expression recognition system that combines active and salient patch based facial emotion identification. The proposed approach is also utilized to extract active patches and some prominent patches from a picture after isolating human face, then facial landmark features. To classify the expression, we computed local binary patterns and discrete cosine transform features.

2. Related work

Human emotions and intents are expressed by facial expressions, and the FER system's key component is determining an effective and efficient feature [11]. Human behavior that plays a vital part in human relations is depicted via expression. Automated face recognition can play an essential role in human-machine interaction, as well as in behavioral analysis and healthcare treatments.

The main issue in human to computer interaction is the recognition of human emotional state because human can understand others' emotions and expressions without any problem, but it is a challenging task for a computer to understand human emotions with accuracy [12]. Facial expression recognition problems are directly connected to the deformation of face region [13]. As different people can have different skin tone or age and gender expressing same emotions and have a small difference in the muscles of face region even the most facial muscles that are involved in the building of an expression can be very hard to understand in machine perspective [14]. Our job is not only to identify these expressions but also to increase the accuracy and efficiency of our system as compared to the previous systems. As some expressions are hard to identify because of their similarity in nature [15]. For example, a person crying, and smiling is easily identified by a human but for computer it is not an easy job to do because of the minor change in facial muscles [16]. So, we must come up with a system that will discriminate and identify these expressions in depth.

In [17], the author used the technique in which first images are divided into the small regions and then local binary patterns are applied on the image regions to extract the appearance of weighted sub regions. Texture features like LBP are effective to use in expression analysis and recognition. In the proposed work we also utilized the LBP features to extract the salient patches.

In this approach, which is distinct from others, the focus will be on statistically learning the most effective sub-regions. The basic goal of facial expression recognition is to accurately classify various facial expressions. In the field of face expression identification, the Support Vector Machine (SVM) is the most widely used machine learning technique. SVM is used in this paper to classify data. SVM is a binary classifier that uses the one-against-one (OAO) technique to classify multiple classes. Majumder [18] used the Kohonen Self-Organizing Map to recognise facial expressions based on their appearance (KSOM). Appearance features are retrieved from evenly subdivided small blocks using uniform Local binary patterns (LBPs), which are then applied to the entire facial image. After that, a dimensionality reduction technique is used. The PCA is used to reduce the size of the LBP feature vector and eliminate superfluous data that adds to the calculation cost. [19] proposed a self-organizing feature map-based automatic face emotion identification system. The Viola and Jones technique is used first to detect a face in a picture. A composite method for locating pupils was proposed after a human face detection. The usage of SOMs was proposed in the research so the located face image could be rotated, cropped, and facial features could be added. Finally, a multi-layer perceptron (MLP) classification approach was used. the classification of six different face emotions, including neutral expression This proposed method is more complicated and has a higher computing cost. Happy [20] suggested a novel expression recognition technique based on appearance attributes of selected facial patches. The placement of feature points, that are extracted actively during emotion extraction, is important in specific face patches. These active patches are then processed to produce the salient patches. It includes discriminative features for classifying each pair of expressions, resulting in distinct face patches being selected as prominent for different pair of emotion classes. These features are used in a one-versus-one classification approach. All of the patches' visual features are given into a multi-class classifier.

For real-time facial expression recognition, an ensemble method is presented that not only extracts the facial regions in real-time but also compresses data using multiple features. It solves issues of insufficient data and expression unrelated intra-class differences using the Local Pathway network and Attention Module. The datasets used for analysis are FER+ and CK+ [21]. To reduce the redundancy and cost of ensemble training, branching is used instead of a single Deep CNN. Redundancy is thus reduced by varying the branching levels, maintaining the diversity and generalization power. Moreover, the generalization error is reduced too. All these experiments are being done on AffectNet and FER+ datasets using an ensemble with shared resources. The study can be further extended for approaches to overcome catastrophic forgetting in ensembles using shared resources [22].

FER based on static images, standardizing a Deep Face Recognition Net (FaceNet2ExpNet) to handle the small datasets by fine-tuning the datasets CK+, Oulu-CASIA, TFD, and SFEW. Introduction of the two-stage (pre-training and refining) training algorithm with a new distribution function and improved high-level expression semantics. Furthermore, other domains having small datasets also used the same method [23]. VGG-like standard deep architecture for a clearer definition of categorizing children's emotions related to learning using theoretical and psychological frameworks is presented. It produces effective state prediction accuracy including neutral and positive states for FER+ Dataset. For meaningful interpretation of information, the work is extended from multiple sources & working with multi-modal datasets [24].

3. Proposed System Overview

The methodology of proposed system starts by human frontal face image acquisition. The six different expression images are used. Gaussian smoothing filter is performed as a preprocessing step to remove the noise. Then this noise removed image is send into the face detection block. This block gives the resultant face image. After detection of face image, face regions like eyes, eyebrows, mouth and nose is extracted from it. By using this information of facial active regions, active and salient patches are extracted using the landmark points on face. Now these patches are fed into the feature calculation block with labels so that classification can be performed. The calculated features are given into the multiclass SVM as input to classify the expression. The proposed system is given below in Figure 1.

3.1 Active and salient patch extraction

The local patches were recovered from the face image during an expression depending on the role of active eye expressions. The active patches are not always in the same place on the face image. Rather, their position is determined by the placements of facial landmarks [23]. All facial patches were kept the same size and were about one-ninth the width of the face. From here on, we'll refer to the patches by the numbers that have been assigned to them. P1, P4, P18, and P19 were retrieved directly from the positions of the lip corners and inner brows, as illustrated in Figure 1.

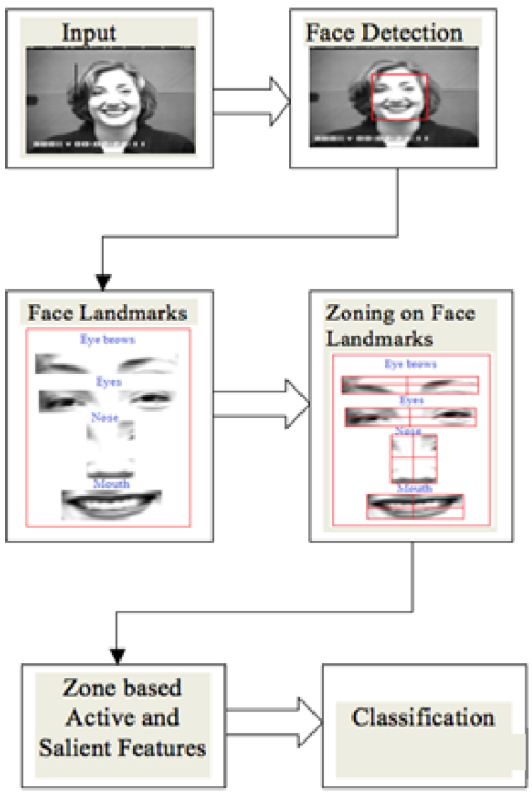

Figure 1: Proposed FER System

P16 was the patch above P16, and it was located in the centre of both eyes. P3 and P6 were found in the middle of the eye and nose. P14 and P15 were just below the eyes. P2, P7, and P8 were grouped together and positioned on one side of the nose. P9 was situated directly beneath P1. P5, P11, P12, and P13 were discovered in a similar manner. P10 was positioned in the middle of P9 and P11. Figure 2 depicts the position of the active facial patch on the frontal face.

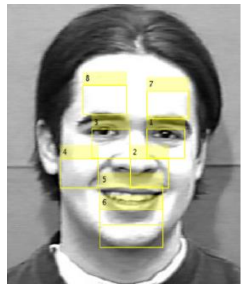

Figure 2: Active Patches Position on the image

According to human perception, not all facial patches are able to recognize a single expression. Each of the face patches that recognizes every expression can be used individually to detect that expression. Based on this idea, we looked into the efficiency of each facial patch for identifying various expressions. Furthermore, certain expressions have comparable facial muscle movements; hence, When detecting expressions, the features of these patches are redundant. As a result, after identifying the active face patches, we selected the key facial patches essential for classification amongst each pair of basic expressions. Changes in the active face region's local binary pattern histogram image are used to extract salient information.

3.2 Feature Extraction

The most crucial aspect of the expression recognition system is feature computation. The discrete cosines transform (DCT) is an effective transform for extracting accurate face expression information. There are two stages to extracting DCT features. The DCT is applied to the full image in the first step to create DCT coefficients, and then some of the coefficients are selected in the second stage to construct feature vectors. The DCT coefficient matrix has the same dimensions as the input picture [24]. In fact, while the DCT does not reduce data dimension by itself, it compresses the majority of signal information into a tiny percentage of coefficients.

To obtain the frequency coefficient matrix of the same dimension, the DCT is applied to the full image. Low frequencies, medium frequencies, and high frequencies are the three bands (sets) that the DCT coefficients are separated into. Low frequencies are related to lighting conditions, while high frequencies are associated with noise and tiny fluctuations. The middle frequencies coefficients include useful information and help to build the image's basic structure. According to the description above, the middle frequency coefficients appear to be better candidates for expression recognition.

So, adding the zones of each feature can increase the computation but it'll also increase the focus of the model on that specific area (feature). Contribution for the specific feature in the model will increase which may give us better results than our base paper.

3.3 Dataset description

The CK+ dataset contains eight expressions (i.e. 7 primary expressions in addition with contempt) that are posed and recorded for more than 200 humans whose age is from 18 years to 50 years respectively. It commonly comprised of Afro-American and Euro-American individuals [25].

For each of the six basic emotions, the Cohn-Kanade collection comprises both male and female face expression image sequences. In this dataset, images were taken in the sequence of images in a certain time period in which first image showed neutral expression that leads to the peak expression of desired category. The images were saved as grayscale and some in color that has size of 640x480 pixels. The dataset consists of 329 peak expression images and 123 neutral images. Neutral and Contempt expression is excluded in this work. Six different expressions i.e. Disgust, Happy, Angry, Fear, Sad and Surprise are classified in this work.

In our tests, we chose the last photograph from each series where the expression was at its most intense. Each expression's number of instances changes depending on its availability. We employed

329 photos in total in our trials on the CK+ database: Anger (41), Disgust (45), Fear (53), Happiness (93), Sadness (56), and Surprise (65).

4.Results discussion

The facial expression recognition framework's performance is evaluated on Cohn Kanade (CK+) dataset of facial expression that is publicly available [36]. CK+ dataset comprises of sequence of images that begins with the neutral expression leads to desired peak expressions. To validate the performance of the proposed system, we used only peak expression from each subject. A 10-cross validation is used for complete dataset. CK+ dataset consists of 593 sequences from 123 subjects. In this dataset sequence of images are available with human's neutral expression and it lasts with a peak expression. Label was given to only 327 of the 593 sequences of human facial expressions; this is because these are the only images that is suitable for the prototypic definition. In our work we used two peak expression images for all fear, anger, sad, disgust, happy and surprise expressions.

It is observed from the dataset that sad, anger and fear expression have slight less no. of image sequences in comparison to all other expressions. We accomplish experimental results for six-class facial expression recognition system. As discussed earlier, face detection was applied on all image sequences in addition to scale them to covert the face into a common resolution.

Few expression images in this dataset are relatively tough to differentiate especially for human eyes.. It was more credible to develop an efficient approach, which could differentiate the expressions using active patches and difficult to discriminate the almost similar expressions like angry and sad etc. For this purpose, the salient patches were extracted to accurately classify similar expressions.
The confusion matrix of six expressions derived using the suggested method. As observed in Table 1, more than 95% accuracy is achieved for all of the six expressions. However, happy expression gave promising results as compared to other expressions. In our opinion, this is because the happy expressions can be categorized by growing corner of mouth and contraction of eyes simply. Except that SVM classifier misclassifies the 3% fear expression as angry and few angry expression images as sad. The reason of such misclassification is that anger and fear expressions contain similar and indirect facial actions.

Table 1: Active Patches Position on the image

	Happy	Surprise	Fear	Angry	Sad	Disgust
Happy	99	0	1	0	0	0
Surprise	5.	94.30	0	0	0	0
Fear	0	0	97	3	0	0
Angry	0	0	0	99.7	0.3	0
Sad	0	0	0	1.9	98.1	0
Disgust	0	0	0	0	3.5	96.5

Table 2 illustrates the performance comparison of different methods and recognition rate of proposed method wit state of art methods. It is noticed that we achieved better accuracy on angry, sad and happy expressions. Few angers expression is misclassified as ad and fear expression as angry. The reason was little change in facial salient areas.   Moreover, the average accuracy is the highest among these methods. The performance doesn't achieve 99%, this is because the image sequences of CK+ dataset are taken in the unrestrained situation (or are real-world images) and are more challenging to classify as compared to other datasets.

Table 2: Recognition rate on CK+ under different approaches

Approach	Liu2015 [33] (LBP)	Ai Sun, 2018 [32] Optimized active area	Sharma, 2019 [31] (Texture feature)	Proposed Method
Classifier	SVM	CNN	CNN	SVM
Recognition rate	92.67	97	97.3	97.48

5.Conclusion and Future work

In this work, we aim to design a robust technique of automatic facial expression analysis and recognition using zone based active and salient patches. The methodology of proposed system starts by human frontal face image acquisition. The six different expression images are used. Face detection is done using the Voila Jones method, and then features are retrieved from facial patches. The active facial patches are found on the areas of the face that alter dramatically during various expressions. After face landmarks are detected, active patches are discovered, and hybrid characteristics are determined from these patches. The computational cost of extracting features is reduced by using small sections of the face rather than the entire face. This block gives the resultant face image, the active facial regions of the face image and zoning is done on it. Based on classification expression is recognized. We evaluated our algorithm extended Cohn-Kanade (CK+) dataset.

In future study, we plan to expand on the proposed model by adding other expressions such as neutral and contempt, as well as evaluating the system's performance in real-time applications. To examine and assess the system's performance, more classifiers can be utilized.

References

[1] HN, Naveen Kumar, S. Jagadeesha, and Amith K. Jain. "Human Facial Expression Recognition from static images using shape and appearance feature." 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT). IEEE, 2016.

Revealed, Emotions. "Recognizing faces and fee lings to improve communication and emotional life." *Henry Holt & Co* (2003).

[2] Calder, Andrew J., and Andrew W. Young. "Understanding the recognition of facial identity and facial expression." *Nature Reviews Neuroscience* 6.8 (2005): 641.

[3] Talele, Kiran, et al. "Facial expression recognition using general regression neural network." *2016 IEEE Bombay Section Symposium (IBSS)*. IEEE, 2016.

[4] Lyons, Michael J., Julien Budynek, and Shigeru Akamatsu. "Automatic classification of single facial images." *IEEE transactions on pattern analysis and machine intelligence* 21.12 (1999): 1357-1362.

[5] Happy, S. L., Anjith George, and Aurobinda Routray. "A real time facial expression classification system using local binary patterns." *2012 4th International conference on intelligent human computer interaction (IHCI)*. IEEE, 2012.

[6] Hadid, Abdenour, Matti Pietikainen, and Timo Ahonen. "A discriminative feature space for detecting and recognizing faces." *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*. Vol. 2. IEEE, 2004.

[7] Zhao, Guoying, and Matti Pietikainen. "Dynamic texture recognition using local binary patterns with an application to facial expressions." *IEEE Transactions on Pattern Analysis & Machine Intelligence* 6 (2007): 915-928.

[8] Kaulard, Kathrin, et al. "The MPI facial expression database—a validated database of emotional and conversational facial expressions." *PloS one* 7.3 (2012): e32321.

[9] Benitez-Quiroz, C.F.; Srinivasan, R.; Martinez, A.M. EmotioNet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 5562–5570.

[10] Tao, S.Y.; Martinez, A.M. Compound facial expressions of emotion. Natl. Acad. Sci. 2014, 111, E1454–E1462.

[11] S. Happy and A. Routray, "Automatic facial expression recognition using features of salient facial patches," *IEEE transactions on Affective Computing*, vol. 6, pp. 1-12, 2015.

[12] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, pp. 1499-1503, 2016.

[13] Z. Yu and C. Zhang, "Image based static facial expression recognition with multiple deep network learning," in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, 2015, pp. 435-442.

[14] A. Iosifidis, A. Tefas, and I. Pitas, "Class-specific reference discriminant analysis with application in human behavior analysis," *IEEE Transactions on Human-Machine Systems*, vol. 45, pp. 315-326, 2015.

[15] S. Happy and A. Routray, "Robust facial expression classification using shape and appearance features," in *Advances in Pattern Recognition (ICAPR), 2015 Eighth International Conference on*, 2015, pp. 1-5.

[16] A. T. Lopes, E. de Aguiar, and T. Oliveira-Santos, "A facial expression recognition system using convolutional networks," in *Graphics, Patterns and Images (SIBGRAPI), 2015 28th SIBGRAPI Conference on*, 2015, pp. 273-280.

[17] K. Zhao, W.-S. Chu, F. De la Torre, J. F. Cohn, and H. Zhang, "Joint patch and multi-label learning for facial action unit and holistic expression recognition," *IEEE Transactions on Image Processing*, vol. 25, pp. 3931-3946, 2016.

[18] P. Liu, S. Han, Z. Meng, and Y. Tong, "Facial expression recognition via a boosted deep belief network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1805-1812.

[19] L. Zhong, Q. Liu, P. Yang, J. Huang, and D. N. Metaxas, "Learning multiscale active facial patches for expression analysis," *IEEE transactions on cybernetics*, vol. 45, pp. 1499-1510, 2015.

[20] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*, 2016, pp. 1-10.

[21] H. Siqueira, S. Magg, and S. Wermter, "Efficient facial feature learning with wide ensemble-based convolutional neural networks," in *Proceedings of the AAAI conference on artificial intelligence*, 2020, vol. 34, no. 04, pp. 5800–5809.

[22] L. Wang, Z. He, B. Meng, K. Liu, Q. Dou, and X. Yang, "Two-pathway attention network for real-time facial expression recognition," *J. Real-Time Image Process.*, vol. 18, no. 4, pp. 1173–1182, 2021.

[23] H. Ding, S. K. Zhou, and R. Chellappa, "Facenet2expnet: Regularizing a deep face recognition net for expression recognition," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, 2017, pp. 118–126

[24] A. H. Farzaneh, Y. Kim, M. Zhou, and X. Qi, "Developing a deep learning-based affect recognition system for young children," in *International Conference on Artificial Intelligence in Education*, 2019, pp. 73–78.

[25] S. Li, J. Xing, Z. Niu, S. Shan, and S. Yan, "Shape driven kernel adaptation in convolutional neural network for robust facial traits recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 222-230.

[26] N. S. Rani, S. K. Verma, and A. Joseph, "A Zone Based Approach for Classification and

Recognition Of Telugu Handwritten Characters," *International Journal of Electrical and Computer Engineering,* vol. 6, p. 1647, 2016.

[27] Y. Liu, Y. Cao, Y. Li, M. Liu, R. Song, Y. Wang, *et al.*, "Facial expression recognition with PCA and LBP features extracting from active facial patches," in *Real-time Computing and Robotics (RCAR), IEEE International Conference on*, 2016, pp. 368-373.

[28] M. Rahul, N. Kohli, and R. Agarwal, "Facial Expression Recognition using Moments Invariants and Modified Hidden Markov Model," *International Journal of Applied Engineering Research,* vol. 13, pp. 6081-6088, 2018.