

Extended LBP based Facial Expression Recognition System for Adaptive AI Agent Behaviour

Kamlesh Mistry

Department of Computer and
Information Sciences
Northumbria University
Newcastle, UK
k.mistry@northumbria.ac.uk

Jyoti Jasekar

IM&T Operations Assistant
NEP Shared System Group
Northumbria Healthcare NHS
Stokesley, UK
jyoti.jasekar@gmail.com

Biju Issac

School of Computing, Media
and the Arts
Teesside University
Middlesbrough, UK
b.issac@tees.ac.uk

Li Zhang

Department of Computer and
Information Sciences
Northumbria University
Newcastle, UK
li.zhang@northumbria.ac.uk

Abstract — Automatic facial expression recognition is widely used for various applications such as health care, surveillance and human-robot interaction. In this paper, we present a novel system which employs automatic facial emotion recognition technique for adaptive AI agent behaviour. The proposed system is equipped with kirsch operator based local binary patterns for feature extraction and diverse classifiers for emotion recognition. First, we nominate a novel variant of the local binary pattern (LBP) for feature extraction to deal with illumination changes, scaling and rotation variations. The features extracted are then used as input to the classifier for recognizing seven emotions. The detected emotion is then used to enhance the behaviour selection of the artificial intelligence (AI) agents in a shooter game. The proposed system is evaluated with multiple facial expression datasets and outperformed other state-of-the-art models by a significant margin.

Keywords — feature optimization, facial expression recognition, local binary pattern, adaptive agent behaviour

I. INTRODUCTION

Facial expression recognition plays an important role in computer vision and human-computer interaction (HCI). Recent studies show that facial expressions are widely used in healthcare [1], video games [2], surveillance systems [3], and humanoid robots [4],[5]. Accurate facial emotion recognition can significantly improve the experience of video game users i.e. the AI agent's behaviour will adapt to the user's emotional state. However, detecting real-time facial emotions is very challenging and is affected by various pose variations, illumination changes, occlusion and noise in the image. Another difficult task is to select significant discriminating facial features that could represent the characteristics of each expression because of the subtlety and variation of facial expressions. Moreover, the current AI agent's behaviour system does not consider any emotional or actual information from the user except for inputs from the video game controller, which leads to predictable AI behaviour.

In order to deal with such challenging real-time tasks, an optimal, robust and accurate facial emotion analysis system is required which can be used in an adaptive AI agent behaviour system. Over the years, many researchers have

proposed mode-specific and parametric feature extraction models to overcome the above-mentioned challenges. Yet, most of the models find it difficult to overcome all the challenges while preserving the high-quality features and low computational complexity. The widely used feature extraction algorithms are Local binary pattern (LBP) [6], Gabor filter [7], and Scale Invariant Feature Transform (SIFT) [8]. These feature extraction models come with their own benefits and risks e.g. LBP is robust to illumination and scaling changes but does not work well with rotation variations [4], Gabor filter is robust to the rotation and illumination changes but works poor with scaling [7]. SIFT shows poor performance when images have illumination and rotation variations [6],[4]. On top of that, none of the above algorithms has been used to deal with adaptive AI agent behaviour system.

Considering the above-mentioned challenges, this work presents a novel variant of feature extraction model and adaptive AI agent behaviour system. The overall system shown in figure 1 consists of three steps, namely feature extraction, emotion classification and adaptive AI agent behaviour with the following major contributions:

- To overcome the challenges such as illumination variations, pose variation and scaling difference, a novel variant of LBP is proposed. The suggested model employs threshold based upper and lower binary patterns to extract ideal discriminative features.
- The overall system is integrated with a shooter game AI agent to gain adaptive AI behaviour in real-time.

The rest of the paper is organized as follows. Section 2 presents related research. Section 3 introduces the proposed adaptive AI agent behaviour system including a novel variant of LBP for feature extraction and emotion recognition. Section 4 presents an evaluation of the proposed system in comparison with other related research. Section 5 draws conclusions and identifies future directions of research.

II. RELATED WORKS

The LBP is an uncomplicated yet very streamlined operator which thresholds the neighbouring pixels and views the resultant pixel as a binary digit [6]. The local structures of

images are aptly summarized by the LBP as it is a non-parametric descriptor [6] and as the LBP operators work in sub-regions of 3x3 pixels of an image. The advantages of LBP include tolerance against changes in grey level illumination and their aforementioned computational simplicity. Their limitations arise in their inability to capture dominant features within large-scale structures through the small 3x3 pixels neighbourhood. Some of the features of LBP, which were originally designed for feature extraction, have been found to be locally effective for face analysis. LBP has proven validity against expression and pose variations and is also insensitive to grey level differences due to illumination discrepancies. Literature shows that numerous variants of LBP have been proposed to deal with the limitations mentioned above. Sharadmani et al. proposed a gradient derivative LBP in [9] for face detection that is virtually homogenous to the consummated LBP magnitude. A novel LBP variant introduced by Pavel et al. in [10] was equipped with an operator that supervised more pixels and diverse neighbourhoods to calculate the feature vector

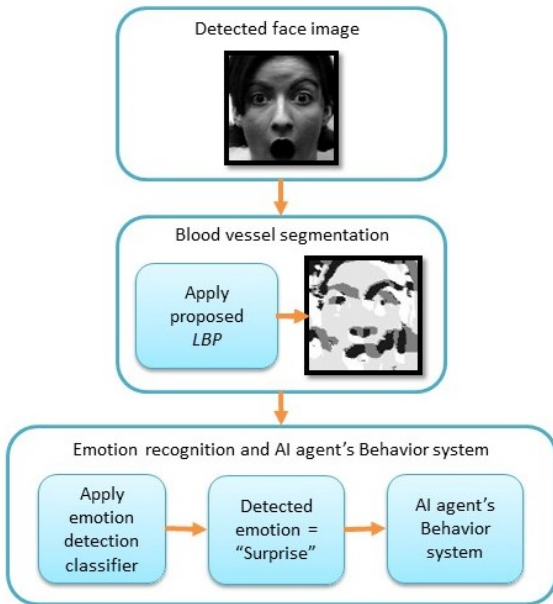


Fig 1. System architecture of the proposed system

Feature extraction methods employing local binary patterns tend to use LBP histograms generated within a rectangular periphery [11]. The joint series of histogram values generate face representation vectors, which are then subjected to comparisons administered by metrics such as histogram intersection or Chi-square distance [12]. Ojala et al. [6] proposed an unorthodox uniform LBP variant, which availed diminished histogram sizes of 59 by classifying rare occurrences within a single value. DTLBP, which is a dynamic threshold LBP extension suggested by Li et al., utilized the average value of the neighbourhood pixels coupled with the maximum contrast amongst the neighbouring points to discover the feature vector [13]. A noise invariant local ternary pattern (LTP) was proposed in [14], which functions in three states to identify and extract distinctions between the central pixel and its neighbourhood.

Zhang et al. [15] proposed a local derivative pattern (LDP) that puts to use high-end features, effectively enabling it to represent more information. The weighted and adaptive LBP based texture descriptor implemented in [16] tackles complexities like insensitiveness to scaling, rotation, viewpoint variations and non-rigid deformations. Structural information within the images can be retained more successfully by instilling an elliptical neighbourhood rather than a circular one as demonstrated by the Elongated Binary Pattern (ELBP) in [17].

The Improved Local Binary Pattern (ILBP) extension proposed by Jin et al. in [18] weighs the intensities of the neighbourhood pixels against the local mean pixel intensity, efficiently reducing noise disturbance. Extended local binary pattern (ELBP) introduced in [19] features two complementary characteristics (pixel intensity and differences) and has tested superior to other classic approaches dealing with texture classification. Although most variants indulge in tweaks to the LBP operator for overall modification, the core schemes for matching and generating feature vectors remain unaltered. A variant that modifies this core mode of operation was proposed in [20] that automatically extracts the distinctive facial features courtesy of Gabor wavelets and K means clustering algorithm. By producing feature vectors in the locations within the frame and comparing them individually, experimental results were obtained that outperformed several classical alternatives. Ojala et al. [21] discussed on the multiresolution grey-scale and rotation invariant texture classification with local binary patterns and Shan et al. [22] explained on the facial expression recognition based on local binary patterns, which are also works that authors noted.

Zhang et al., propose a novel deep learning framework, called spatial-temporal recurrent neural network (STRNN) for emotion recognition. To capture those spatially co-occurrent variations of human emotions, a multidirectional recurrent neural network (RNN) layer is employed to capture long-range contextual cues by traversing the spatial regions of each temporal slice. A bi-directional temporal RNN layer is further used to learn the discriminative features characterizing the temporal dependencies of the sequences. The experimental results show that the proposed method is better than state-of-the-art methods [23]. Zhang et al. proposed a new emotion recognition system based on facial expression images by enrolling 20 subjects and let each subject pose seven different emotions: happy, sadness, surprise, anger, disgust, fear, and neutral. They used biorthogonal wavelet entropy to extract multiscale features and used fuzzy multiclass support vector machine to be the classifier showing it is better than three state-of-the-art methods [24].

Shojaeilangari et al. proposed extreme sparse learning, which can jointly learn a dictionary and a nonlinear classification model. The proposed approach combines the discriminative power of extreme learning machine with the reconstruction property of sparse representation to enable accurate classification when presented with noisy signals and imperfect data recorded in natural settings, which eventually enables to achieve the state-of-the-art recognition

accuracy [25]. Zen et al. proposed a regression framework which exploits auxiliary annotated data to learn the relation between person-specific sample distributions and parameters of the corresponding classifiers. Then, when considering a new target user, the classification model is computed by simply feeding the associated sample distribution into the learned regression function. They evaluated the approach in different applications like pain recognition and action unit detection using visual data and gestures classification using inertial measurements [26].

III. THE PROPOSED SYSTEM

In this section we introduce the proposed adaptive AI agent behaviour system. The overall system consists of two key steps, i.e. a novel variant of LBP-based feature extraction, and emotion recognition. Each step is introduced in detail in the following.

A. Feature Extraction

In feature extraction process, pre-processing is applied to reduce image noise. A histogram equalization method is initially used to improve the contrast of an input image. Then, a bilateral filter is applied to reduce image noise, while preserving the edges. We subsequently apply Viola and Jones's face detection algorithm provided in the OpenCV package to detect the face region of the input image. The detected face is further processed using *proposed LBP* to extract robust features.

1) The Conventional LBP

Ojala et al. [6] proposed the conventional LBP which thresholds each of the 3x3 neighbouring pixels with a centre pixel value. The conventional LBP was further extended to use various numbers of circular neighbouring pixels [27]. The LBP operator $LBP_{p,r}$ can produce 2^p different binary patterns, where p denotes the number of neighborhood pixels and r denotes the radius of the circular pattern. The equation for calculating the $LBP_{p,r}$ operator can be given as follows:

$$LBP_{p,r} = \sum_{p=0}^{p-1} S(g_p - g_c) 2^p, S(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where g_p denotes the neighbourhood pixel at location p and g_c is the centre pixel. The important information such as edges, corners, spot and the flat area can be detected using the LBP [27]. The conventional LBP is robust to illumination and scaling variations but fails to deal with rotation variations [28]. Whereas, the gradient images contain enhanced edge information and are more stable than raw pixel intensities, which can benefit to deal with rotation and illumination variations.

2) The Proposed Variant of LBP with Kirsch Operator

The proposed system is similar to the existing variants of LBP i.e. overlap-LGBP [29] and hvn-LGBP [30]. The major difference between the proposed LBP and the above variants is the choice of the convolution filter. The variants employ Gabor filters which are calculated by a continuous function that can be scaled and rotated to any subjective orientation. The construction of Gabor kernels requires high

computation efforts and modulation of Gaussian function by a complex plane wave. The Kirsch operator or Kirsch compass kernel [31] is a non-linear edge detector and it is used to find the maximum edge strength in a few predetermined directions.

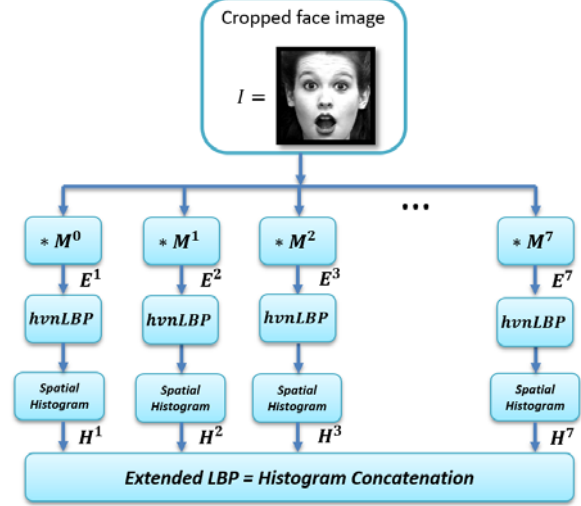


Fig 2. The comparison of original LBP and *proposed LBP*

The Kirsch operator uses fixed kernels oriented at 45-degree intervals, which can be implemented with less computational efforts. The proposed LBP model employs Kirsch compass masks which are set to fixed kernel orientation at 45-degree intervals. Compared to the Gabor-based models, the proposed LBP model requires less computational time while maintaining the quality of the extracted features. First, the original image is processed with a Kirsch operator to generate eight edge responses through all 8 compass directions: N, NW, W, SW, S, SE, E, and NE. Then, each edge response is treated separately to exploit the spatial relationships among neighbours within that edge response. The edge magnitude of the Kirsch operator is calculated as the maximum magnitude across all the directions:

$$h_{n,m} = \max_{z=1,2,\dots,8} \sum_{i=-1}^1 \sum_{j=-1}^1 g_{ij}^{(z)} \cdot f_{n+i,m+j} \quad (2)$$

where z enumerates the compass direction kernels g as follows.

$$\mathbf{g}^{(1)} = \begin{bmatrix} +5 & +5 & +5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}, \mathbf{g}^{(2)} = \begin{bmatrix} +5 & +5 & -3 \\ +5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix},$$

$$\mathbf{g}^{(3)} = \begin{bmatrix} +5 & -3 & -3 \\ +5 & 0 & -3 \\ +5 & -3 & -3 \end{bmatrix}, \mathbf{g}^{(4)} = \begin{bmatrix} -3 & -3 & -3 \\ +5 & 0 & -3 \\ +5 & +5 & -3 \end{bmatrix} \text{ etc.}$$

The mathematical representation of the proposed LBP model is given below:

$$E^i = I \times M^i, \text{ where } i = 0, 1, 2, 3, \dots, 7 \quad (3)$$

where I is the original image, E^i denotes the eight edge responses computed using kirsch operator M^i . Then the

horizontal and vertical neighbourhood LBP (*hvnLBP*) is applied on each of the edge responses to calculate $hvnLBP_{p,r}$.

$$hvdnLBP_{p,r} = \{S(\max(e_{p_0}^i, e_{p_1}^i, e_{p_2}^i)), S(\max(e_{p_7}^i, e_{p_3}^i)), S(\max(e_{p_6}^i, e_{p_5}^i, e_{p_4}^i)), S(\max(e_{p_0}^i, e_{p_7}^i, e_{p_6}^i)), S(\max(e_{p_1}^i, e_{p_5}^i)), S(\max(e_{p_2}^i, e_{p_3}^i, e_{p_4}^i))\} \quad (4)$$

where $e_{p_i}^i$ denotes the edge response values of the neighborhood pixels in i^{th} edge response E^i , r is the radius, and S denotes the comparison operation, as follows.

$$S(\max(e_{p_j}^i, e_{p_k}^i, e_{p_m}^i)) = \begin{cases} 1 & \text{if maximum} \\ 0 & \text{if not maximum} \end{cases} \quad (5)$$

where, $e_{p_j}^i$, $e_{p_k}^i$ and $e_{p_m}^i$ denotes the edge response values of the neighborhood pixels in a row and column. Note that $e_{p_k}^i$ is removed if it represents the center pixel. The Figure 2 illustrates the proposed facial feature extraction mechanism. In comparison to conventional LBP, the proposed extended *hvnLBP* operator captures more discriminative contrast information and can achieve better face representation. The Figure 3 shows the output results generated using proposed LBP variant for different illumination variations. The illuminated images in Figure 3 are generated using Adobe Photoshop tool to show proposed LBP's robustness to illumination variations. Even with illumination variations the proposed LBP can, generate very identical outputs. Due to its discriminative capability, the proposed LBP can show a promising representation of the important features of the face.

In this work, the resolution of the detected face image is 75×75 pixels and after applying the proposed *LBP* operator, the face image is divided into 25×25 (i.e. 625) sub-regions with the size of each sub-region being 3×3 . In comparison to conventional LBP and its variants, the proposed *LBP* operator captures more discriminative contrast information such as corners and edges among the neighbourhoods. The feature extracted by proposed LBP are further used as input features to train and test the various classifiers.

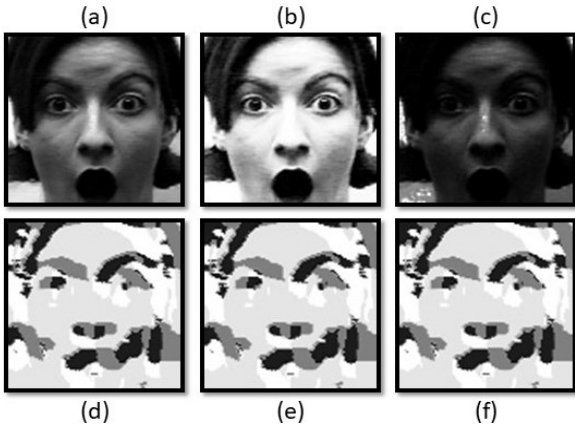


Fig 3. (a) Original illumination image, (b) bright illumination image, (c) low light illumination image, and (d), (e), (f) are the proposed LBP generated outputs.

B. Emotion Classification

In this work, we have employed diverse classifiers to detect the seven basic emotions (i.e. anger, happiness, sadness, surprise, disgust, fear, and neutral). NN, multi-class SVM, and the SVM-based and NN-based ensembles with SVM and NN as base classifiers respectively are employed to conduct emotion classification. The selected features retrieved by the mGA embedded FA are used as inputs to the classifiers. The input layer nodes for NN is set to a number of features extracted from proposed LBP, where each node indicates the feature extracted by proposed LBP. The NN classifier has one hidden layer and one output layer with seven nodes representing each emotion respectively. Moreover, the grid-search method is also employed to obtain the most favourable parameter settings for the SVM classifier to achieve optimal performance. The optimal settings obtained for each single model NN and SVM mentioned above are also applied to the setting of each base classifier within each ensemble. Both ensembles employ three base classifiers and a weighted majority voting combination method to produce final classification.

We had employed a set of 1250 images in total from the CK+ [32], MMI [33], JAFFE [34], Bosphorus 3D [35] and BU-3DFE [36] database for training all the classifiers. A test set of 200 images is extracted respectively from each of the CK+, MMI, JAFFE, Bosphorus 3D and BU-3DFE databases for system evaluation. Overall, the NN and SVM-based ensembles achieve the best accuracy when tested with images from the five databases.

IV. EVALUATION

The proposed system is tested by integrating with a shooter game's AI agent. The novel facial emotion recognizer is implemented from scratch and combined with AI agent's behaviour system to enable a real-time adaptive behaviour for the AI agent. The evaluation of the system usually starts with a basic behaviour tree, which has four different behaviours (shoot, take cover, wander and flee). To control the behaviour using facial expressions, we have introduced a threshold value for each behaviour. These threshold values for each behaviour will change depending on the user's facial expression. For example, if the user's facial expression is disgust or surprise, the AI agent will behave more aggressively by increasing the threshold for shoot behaviour. Table 1 shows the facial expression and its associated behaviours.

TABLE 1. FACIAL EXPRESSIONS AND BEHAVIOUR'S ASSOCIATED WITH IT

	Happy and Neutral	Angry, fear and Sadness	Disgust and surprise
Shoot	0.50	0.10	1.00
Take cover	1.00	0.50	0.50
Wander	0.50	1.00	0.50
Flee	0.10	1.00	0.10

First, we show the comparison between proposed LBP and state-of-the-art LBP variants. Then mGA embedded FA-based feature selection algorithm is evaluated against the conventional FA, GA, PSO and FA variants. Single and ensemble classifiers such as NN, SVM, and NN-based and

SVM-based ensembles, are used for the classification of seven emotions. For all the experiments, a set of 1250 images from CK+, JAFFE, MMI, BU-3DFE and Bosphorus 3D is employed for training while a set of 200 images from each of five datasets (i.e. CK+, JAFFE, MMI, BU-3DFE and Bosphorus 3D) are used for testing.

The first experiment conducted used 1250 and 1000 images from five datasets for training and testing, respectively. Table 2 shows the results of the proposed LBP against state-of-the-art feature extraction algorithms (i.e. PCA, SIFT, LBP, LGBP, CS-LBP, LDP, LTP and DTLBP). In order to conduct the comparison between the proposed LBP and other texture descriptors, table 2 presents the results obtained using all the raw features extracted by the proposed LBP and other texture descriptors without any feature selection method.

TABLE 2. COMPARISON BETWEEN THE LBP OPERATOR AND OTHER TEXTURE DESCRIPTORS

	NN (%)	SVM (%)	NN-based Ensemble (%)	SVM-based Ensemble (%)
PCA	66.82	67.87	69.21	70.15
SIFT	69.80	70.54	72.64	72.75
LBP	69.83	69.50	72.18	72.21
LGBP	70.87	71.64	73.51	74.00
CS-LBP	74.34	73.91	75.91	76.45
LDP	75.33	76.49	78.00	78.93
LTP	77.54	77.74	80.36	80.66
DTLBP	75.84	75.79	78.46	78.50
Proposed LBP	81.56	82.50	87.24	89.30

For each algorithm, results obtained by the SVM-based ensemble classifier outperform those of all the other classifiers. The highest accuracy achieved by the proposed LBP is 89.30% with the SVM-based ensemble and outperforms PCA, SIFT, LBP, LGBP, CS-LBP, LDP, LTP, and DTLBP by 19.15%, 16.55%, 17.09%, 15.30%, 12.85%, 10.37%, 8.64%, and 10.80%, respectively. The results gained using the proposed LBP show significant improvement over those obtained using other texture descriptors, which proves the efficiency of the proposed LBP-based feature extraction algorithm.

In order to further demonstrate the efficiency of the proposed system, we have conducted separate experiments using four different classifiers (i.e. NN, SVM, NN-based ensemble, and SVM-based ensemble) for each dataset, respectively. Table 3 shows the comparison results for four classifiers when evaluated with CK+ database.

As shown in Table 3, the proposed LBP in combination with ensemble-SVM outperforms the rest of the algorithms with a significant margin when evaluated with CK+ datasets. In Tables 4, 5, 6, and 7 we have compared the performance of the proposed LBP with PCA, SIFT, LBP, LGBP, CS-LBP, LDP, LTP, and DTLBP when evaluated with MMI, JAFFE, BU-3DFE and Bosphorus-3D datasets. As illustrated in Tables 4, 5, 6, and 7, the proposed LBP variant achieves the highest accuracy for all four classifiers, outperforming rest of the selected algorithms by a significant margin, respectively.

TABLE 3. COMPARISON BETWEEN THE LBP OPERATOR AND OTHER TEXTURE DESCRIPTORS (CK+ DATASET)

	NN (%)	SVM (%)	NN-based Ensemble (%)	SVM-based Ensemble (%)
PCA	85.99	88.89	91.11	92.97
SIFT	87.45	89.57	91.70	93.67
LBP	88.74	90.41	93.21	94.26
LGBP	92.16	91.10	93.19	94.65
CS-LBP	90.79	92.98	95.42	97.13
LDP	92.38	93.94	96.52	98.02
LTP	94.44	95.46	98.15	99.81
DTLBP	93.15	93.70	97.88	98.10
Proposed LBP	97.33	98.82	100	100

TABLE 4. COMPARISON BETWEEN THE PROPOSED LBP OPERATOR AND OTHER TEXTURE DESCRIPTORS (MMI DATASET)

	NN (%)	SVM (%)	NN-based Ensemble (%)	SVM-based Ensemble (%)
PCA	81.73	84.42	86.62	88.04
SIFT	85.08	86.76	89.76	90.85
LBP	85.48	86.07	88.43	89.64
LGBP	85.95	87.60	89.72	90.94
CS-LBP	86.02	89.49	92.27	94.26
LDP	87.79	90.74	93.38	94.38
LTP	88.98	84.42	92.65	93.88
DTLBP	88.67	85.31	92.46	94.00
Proposed LBP	92.45	94.04	97.90	98.30

TABLE 5. COMPARISON BETWEEN THE PROPOSED LBP OPERATOR AND OTHER TEXTURE DESCRIPTORS (JAFFE DATASET)

	NN (%)	SVM (%)	NN-based Ensemble (%)	SVM-based Ensemble (%)
PCA	85.65	86.63	89.36	90.38
SIFT	89.01	89.36	92.33	93.36
LBP	89.19	88.57	90.61	92.50
LGBP	90.76	89.65	92.24	93.36
CS-LBP	90.57	92.44	94.65	96.19
LDP	90.54	93.97	96.96	98.64
LTP	93.84	93.47	95.52	96.60
DTLBP	92.74	93.63	94.66	95.38
Proposed LBP	96.21	97.36	100	100

TABLE 6. COMPARISON BETWEEN THE PROPOSED LBP OPERATOR AND OTHER TEXTURE DESCRIPTORS (BU-3DFE DATABASE)

	NN (%)	SVM (%)	NN-based Ensemble (%)	SVM-based Ensemble (%)
PCA	81.70	83.02	85.03	86.35
SIFT	84.50	84.61	87.39	88.91
LBP	85.14	86.66	89.28	90.41
LGBP	85.72	86.22	88.76	90.52
CS-LBP	84.76	87.53	90.26	91.47
LDP	87.80	89.71	92.02	93.98
LTP	88.24	89.50	92.18	93.20
DTLBP	88.45	90.10	93.26	94.73
Proposed LBP	91.55	93.15	98.11	98.75

TABLE 7. COMPARISON BETWEEN THE PROPOSED LBP OPERATOR AND OTHER TEXTURE DESCRIPTORS (BOSPHORUS-3D DATASET)

	NN (%)	SVM (%)	NN-based Ensemble (%)	SVM-based Ensemble (%)
PCA	83.06	86.54	89.47	90.66
SIFT	86.52	88.21	90.93	92.64
LBP	86.60	89.44	92.34	93.39
LGBP	88.88	90.94	93.39	94.97
CS-LBP	89.94	91.57	94.05	96.03
LDP	91.11	92.35	94.54	96.53
LTP	91.83	92.02	94.47	95.60
DTLBP	90.44	91.99	93.80	95.00
Proposed LBP	94.60	95.97	97.22	97.33

TABLE 8. COMPARISON WITH RELATED RESEARCH FOR DIFFERENT DATASETS

Methods	Methodology	CK+	MMI	JAFFE	BU-3DFE	Bosphorus 3D
Shan et al. [22]	Boosted LBP+SVM	91.40	86.9	81	-	-
Elaiwat et al. [35]	Spatio-temporal RBM based model	95.66	81.63	-	-	-
Zhong et al. [36]	CSPL	89.89	73.53	-	-	-
Derkach and Sukno [37]	Graph Laplacian	-	-	-	81.5	-
Jan and Meng [38]	ULBP+LPQ+EO H+83P+FD	-	-	-	88.32	79.46
This work	Proposed System	100	98.30	100	98.75	98.02

The evaluation results show that the proposed LBP in combination with an SVM-based ensemble can achieve the highest accuracy for all five datasets. The proposed system is further compared with other existing state-of-the-art facial expression recognition research. Table 8 presents the detailed comparison between the proposed system and other related research for various databases, respectively.

As indicated in Table 8, the proposed system outperforms all other related research when using all five datasets for training and testing. The proposed LBP-based feature extraction account for the efficiency and robustness of the proposed system.

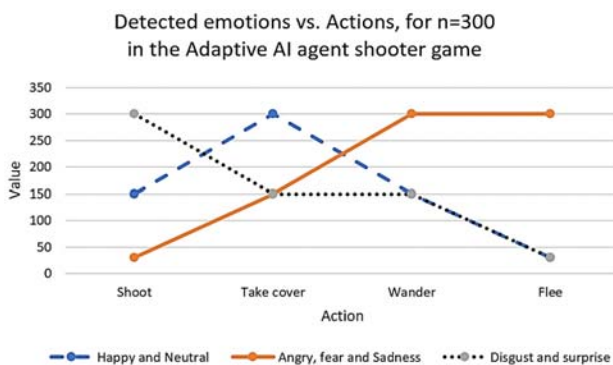


Fig 4. Sample interaction between detected emotions and actions in the shooter game with AI adaptive agent

A sample interaction in the shooter game with the adaptive variable of the number of instances $n=300$, is shown in figure 4 that shows the detected emotions and the corresponding actions involved.

V. CONCLUSION

In this paper we have proposed an adaptive AI agent behaviour system with the proposed LBP variant for feature extraction. Diverse classifiers are employed for recognizing seven facial expressions. The proposed LBP variant can extract discriminative features, which are robust to illumination, scaling and rotation variations. It outperforms other state-of-the-art feature extraction methods such as PCA, SIFT, LBP, LGBP, CS-LBP, LDP, LTP, and DTLBP, significantly. The overall system has been integrated with a shooter game's AI agent to stimulate real-time adaptive AI agent behaviour. The proposed system achieves an average accuracy of 89.30% (using raw features without any feature selection) when evaluated with combined test images from five datasets. The average accuracy of five datasets of LBP variant with the SVM-based ensemble is 99.01%. The system also shows promising performance for each dataset evaluation and achieves an average accuracy of 98.02% for Bosphorus 3D, 98.75% for BU-3DFE, 98.30% for MMI, 100% for both CK+ and JAFFE, respectively. It thus shows a promising performance when compared with other state-of-the-art related facial expression recognition research, thus allowing intelligent interaction in the shooter game. In future work other hybrid or multi-objective feature selection models will also be explored for solving high dimensionality problems.

REFERENCES

- [1] L. Zhang, B. Fielding, P. Kinghorn, and K. Mistry, A Vision Enriched Intelligent Agent with Image Description Generation. In *Proceedings of International Conference of Autonomous Agents and Multiagent Systems*. Singapore, 2016.
- [2] A. S. G'Mussel, and J. Hewig, The value of a smile: Facial expression affects ultimatum-game responses. *Judgment and Decision Making*, 8 (3), 381-385, 2013.
- [3] E. Vural, M. Cetin, A. Ercil, G. Littlewort, M. Bartlett, and J. Movellan, Automated Drowsiness Detection for Improved Driver Safety. In *Proceedings of the International Conference on Automotive Technologies*, 2008.
- [4] L. Zhang, M. Jiang, D. Farid, and A. M. Hossain, Intelligent Facial Emotion Recognition and Semantic-based Topic Detection for a Humanoid Robot. *Expert Systems with Applications*, 40, pp. 5160-5168, 2013.
- [5] L. Zhang, K. Mistry, M. Jiang, S. C. Neoh, and M. A. Hossain, Adaptive facial point detection and emotion recognition for a humanoid robot, *Computer Vision and Image Understanding* 140, 93-114, 2015.
- [6] T. Ojala, M. Pietikäinen, and D. Harwood, A comparative study of texture measures with classification based on featured distribution, *Pattern Recognition* 29 (1), 51-59, 1996.
- [7] M. Valstar and M. Pantic. "Fully automatic facial action unit detection and temporal analysis", in *Proceedings of IEEE Conference on computer vision and pattern recognition (CVPR)*, 2012, pp. 149-156.
- [8] D. G. Lowe, Object recognition from local scale invariant features, In *Proceedings of International conference on computer vision*, pp. 1150-1157, 1999.

- [9] D. Sharadamani and C. NagaRaju, "Face Recognition Using Gradient Derivative Local Binary Patterns", *International Journal of Applied Engineering Research* ISSN 0973-4562 Volume 12, Number 7 pp. 1316-1323, 2017.
- [10] P. Kral' and A. Vrbal, "Enhanced Local Binary Patterns for Automatic Face Recognition", Cornell University Library, arXiv:1702.03349 [cs.CV], February 2017.
- [11] T. Ahonen, A. Hadid, and M. Pietik'ainen, "Face recognition with local binary patterns", in *Computer visioneccv*, pp. 469–481. Springer, 2004.
- [12] D. Zhang and G. Lu, "Evaluation of similarity measurement for image retrieval", in *Neural Networks and Signal Processing, 2003. Proceedings of the 2003 International Conference on*. IEEE, vol. 2, pp. 928–931, 2003.
- [13] W. Li, P. Fu, and L. Zhou, "Face recognition method based on dynamic threshold local binary pattern", in *Proceedings of the 4th International Conference on Internet Multimedia Computing and Service*. ACM, pp. 20–24, 2012.
- [14] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions", *Image Processing, IEEE Transactions on*, vol. 19, no. 6, pp. 1635–1650, 2010.
- [15] B. Zhang, Y. Gao, S. Zhao, and J. Liu, "Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor", *Image Processing, IEEE Transactions on*, vol. 19, no. 2, pp. 533–544, 2010.
- [16] R. Davarzani and S. Mozaffari, "Robust image description with weighted and adaptive local binary pattern features", in *Pattern Recognition (ICPR), 2014 22nd International Conference on*. IEEE, pp. 1097–1102, 2014.
- [17] S. Liao and A. C. S Chung, "Face recognition by using elongated local binary patterns with average maximum distance gradient magnitude", in *Asian conference on computer vision*. Springer, pp. 672–679, 2007.
- [18] H. Jin, Q. Liu, H. Lu, and X. Tong, "Face detection using improved lbp under bayesian framework", in *Image and Graphics (ICIG'04), Third International Conference on*. IEEE, pp. 306–309, 2004.
- [19] L. Liu, L. Zhao, Y. Long, G. Kuang, and P. Fieguth, "Extended local binary patterns for texture classification", *Image and Vision Computing*, vol. 30, no. 2, pp. 86–99, 2012.
- [20] L. Lenc and P. Kr'al, "Local binary pattern based face recognition with automatically detected fiducial points", *Integrated Computer-Aided Engineering*, vol. 23, no. 2, pp. 129–139, March 2016.
- [21] T. Ojala, M. Pietik'ainen, T. M'äenp'ää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (7), 971–987, 2002.
- [22] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: a comprehensive study", *Image Vis. Comput.* 27, 803–816, 2009.
- [23] T. Zhang, W. Zheng, Z. Cui, Y. Zong and Y. Li, "Spatial-Temporal Recurrent Neural Network for Emotion Recognition," in *IEEE Transactions on Cybernetics*, vol. PP, no. 99, pp. 1-9, 2018.
- [24] Y. D. Zhang et al., "Facial Emotion Recognition Based on Biorthogonal Wavelet Entropy, Fuzzy Support Vector Machine, and Stratified Cross Validation," in *IEEE Access*, vol. 4, pp. 8375-8385, 2016.
- [25] S. Shojailangari, W. Y. Yau, K. Nandakumar, J. Li and E. K. Teoh, "Robust Representation and Recognition of Facial Emotions Using Extreme Sparse Learning," in *IEEE Transactions on Image Processing*, vol. 24, no. 7, pp. 2140-2152, July 2015.
- [26] G. Zen, L. Porzi, E. Sangineto, E. Ricci and N. Sebe, "Learning Personalized Models for Facial Expression Analysis and Gesture Recognition," in *IEEE Transactions on Multimedia*, vol. 18, no. 4, pp. 775-788, April 2016.
- [27] S. C. Neoh, L. Zhang, M. Mistry, M. A. Hossain, C. P. Lim, N. Aslam, and P. Kinghorn, "Intelligent Facial Emotion Recognition Using a Layered Encoding Cascade Optimization Model". *Appl Soft Comput.* Volume 34, 72–93, 2015.
- [28] K. Mistry, L. Zhang, S. C. Neoh, C. P. Lim, and B. Fielding, "A micro-GA Embedded PSO Feature Selection Approach to Intelligent Facial Emotion Recognition." *IEEE Transactions on Cybernetics*, 2016. doi: 10.1109/TCYB.2016.2549639
- [29] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended Cohn–Kanade dataset (CK+): a complete expression dataset for action unit and emotion-specified expression, in: *Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis*, San Francisco, USA, pp. 94–101, 2010.
- [30] M. Pantic, M. F. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis, in: *Proceedings of IEEE Int'l Conf. Multimedia and Expo, Amsterdam, The Netherlands*, pp. 317–321, 2005.
- [31] R. Kirsch, "Computer determination of the constituent structure of biological images". *Computers and Biomedical Research*. 4: 315–328, 1971.
- [32] M. Lyons, S. Akamatsu, M. Kamachi and J. Gyoba, "Coding facial expressions with gabor wavelets, in: *Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 200–205, 1998.
- [33] A. Savran, N. Alyuz, H. Dibeklioglu, O. Celiktutan, B. Gokberk, B. Sankur, and L. Akarun, "Bosphorus database for 3D face analysis, *Proceedings of First COST 2101 Workshop on Biometrics and Identity Management*, Denmark, pp. 47–56, 2008.
- [34] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3D facial expression database for facial behavior research, in: *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, pp. 211–216, 2006.
- [35] S. Elaiwat, M. Bennamoun, and F. Boussaid, "A spatio-temporal RBM-based model for facial expression recognition, in: *Pattern Recognition*, 49, pp. 152-161, 2016.
- [36] L. Zhong, Q. Liu, P. Yang, B. Liu, J. Huang, and D. Metaxas, "Learning active facial patches for expression analysis, in: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2562–2569, 2012.
- [37] D. Derkach, and F. M. Sukno, "Local shape spectrum analysis for 3D facial expression recognition, in: *IEEE 12th International Conference on Automatic Face & Gesture Recognition*, 2017.
- [38] A. Jan, and H. Meng, "Automatic 3D facial expression recognition using geometric and texture feature fusion, in: *11th International Conference on Automatic Face & Gesture Recognition*, 2015.