Emotion Recognition for Sentiment Analysis using Basic Feature Extraction Techniques

Introduction to the Application and Its Significance

Facial emotion recognition (FER) is a valuable application in computer vision and affective computing. It seeks to recognize and classify human emotions based on facial expressions in photos or video feeds. The significance of FER rests in its broad applications across numerous domains:

- Human-Computer Interaction: FER can improve the user experience by enabling computers to respond intelligently to users' emotional states, resulting in more intuitive and empathic interfaces.
- Mental Health Monitoring: In psychology and psychiatry, FER systems can aid in the early detection of mood disorders and track emotional well-being over time, potentially revolutionising mental health care.
- Marketing and Customer Service: Real-time analysis of customer reactions to products or services can provide organizations with significant insights into consumer behavior and preferences.
- Education: FER can be used to assess student involvement and emotional responses during learning activities, allowing educators to modify their teaching approaches for improved results.
- Security and surveillance: In high-security settings, FER can help detect suspicious behavior or potential threats based on emotional indicators, hence improving public safety measures.

The FER system built in this study intends to contribute to these applications by creating a reliable and accurate emotion classification model.

Dataset used for creating FER system

- Dataset Overview:
  - The CK+ dataset consists of 593 video clips from 123 participants aged 18 to 50 years.
  - Each video sequence begins with neutral face expression and escalates to the highest representation of one of seven basic emotions: anger, contempt, disgust, fear, happiness, sadness, neutral and surprise.
  - The dataset contains both grayscale and color images, with each frame measuring 640x490 pixels in size.
  - The images have been annotated with locations of 68 facial landmarks, which can be utilized for facial expression analysis and alignment.
  - The dataset is separated into three subsets: training, public test, and private testing.

The CSV file has the following columns:

  - Emotion: Integer value representing emotion label (0: Anger, 1:Disgust, 2: Fear, 3: Happiness, 4: Sadness, 5: Surprise, 6: Neutral, 7: Contempt).
  - pixels: String of pixels values separated by spaces, representing the grayscale pixel values of 48x48 image.
  - usage: String indicating subset image belongs to (Training, PrivateTest, or PublicTest).

- Dataset download location: https://www.kaggle.com/datasets/davilsena/ckdataset

Experimental Setup

1. Dataset:
   - CK+ dataset
   - Dataset includes both grayscale and color images, with each frame being 640x490 pixels in size.
   - Each image is labeled with one of 7 emotion categories: anger, contempt, disgust, fear, happiness, sadness, neutral and surprise.

2. Feature Extraction Methods:
   a) Histogram of Oriented Gradients (HOG)
   b) Local Binary Patterns (LBP)
   c) Facial Landmarks (using dlib's 68-point predictor)
   d) Geometric Features (derived from facial landmarks)
   e) Texture Features (using Gray-Level Co-occurrence Matrix)
   f) Temporal Features (difference between current and previous frame features)

3. Parameters Being Varied:

a) HOG Parameters:

- Window size: (48, 48) and (64, 64)
- Block size: (16, 16) and (32, 32)
- Block stride: (8, 8) and (16, 16)
- Cell size: (8, 8) and (16, 16)
- Number of bins: 9 (fixed)

b) LBP Parameters:

- Radius: 1 and 2
- Number of points: 8 and 16
- Method: 'uniform' (fixed)

c) Texture Parameters:

- Distances: 1 and 2
- Angles: [0, π/4, π/2, 3π/4] (fixed)
- Properties: ['contrast', 'dissimilarity', 'homogeneity', 'energy', 'correlation'] (fixed)

4. Machine Learning Model:

- Random Forest Classifier
- Number of estimators: 100 (fixed)
- Random state: 42 (fixed for reproducibility)

5. Evaluation Metrics:
    - Accuracy
    - F1-score (weighted average)
    - Sensitivity

6. Experimental Procedure:
    a) Load the ck+ dataset and rescale images to 48x48 pixel grayscale.
    b) For each combination of HOG, LBP, and texture parameters:
        - Extract features from all images
        - Split the dataset into 80% training and 20% testing sets
        - Scale the features using StandardScaler
        - Train a Random Forest classifier on the training set
        - Evaluate the model on the test set
        - Record accuracy and F1-score
    c) Analyze results:
        - Identify the best-performing parameter combination
        - List top 5 parameter combinations
        - Perform parameter sensitivity analysis

7. Additional Analyses:
    - Confusion matrix for best-performing model
    - Detailed classification report for best-performing model
8. Visualization:
    - Bar plot comparing accuracy, sensitivity and F1-score across experiments
    - Heatmap of confusion matrix for best-performing model

* Note: Same experimental setup can be used with FER2013 dataset csv (https://huggingface.co/spaces/mxz/emtion/resolve/c697775e0adc35a9cec32bd4d3484b5f5a263748/fer2013.csv)

Description of Selected Feature Extraction Techniques and Implementation Details

The code employs a combination of basic feature extraction techniques to capture various aspects of facial expressions:

a) Histogram of Oriented Gradients (HOG)

b) Local Binary Patterns (LBP)

c) Facial Landmarks

d) Texture Features (GLCM)

e) Geometric Features

f) Temporal Features

a) Histogram of Oriented Gradients (HOG):

- Implementation: Using OpenCV's HOGDescriptor
- Parameters: Window size, block size, block stride, cell size, and number of bins
- Purpose: HOG extracts edge and gradient information from the face image, which is critical for detecting facial features and their orientations.
- Implementation details: The code lets you to experiment with various HOG parameters, including window sizes of (48, 48) and (64, 64).

b) Local Binary Patterns (LBP):

- Implementation: Using skimage's local_binary_pattern function
- Parameters: Radius, number of points, and method (uniform)
- Purpose: It extract texture information from a face image in order to discern between different facial expressions.
- Implementation details: Using uniform technique, system tests LBP at radii of 1 and 2, as well as 8 or 16 points.

c) Facial Landmarks:

- Implementation: Using dlib's face landmark predictor
- Purpose: Technique locates 68 key points on face, which serve as basis for geometric feature extraction.
- Implementation details: Landmarks are extracted using a pre-trained model (shape_predictor_68_face_landmarks.dat).

d) Geometric Features:

- Implementation: Custom function calculating distances and ratios between landmarks
- Features: Eye distance, nose-mouth distance, mouth width, and various facial ratios
- Purpose: These features capture spatial relationships between facial features, which can change with different emotions.

e) Texture Features:

- Implementation: Using skimage's graycomatrix and graycoprops functions
- Parameters: Distances, angles, and properties (contrast, dissimilarity, homogeneity, energy, correlation)
- Purpose: It extract detailed texture information from the Gray-Level Co-occurrence Matrix, providing additional context for emotion classification.
- Implementation details: Code experiments with different distances for texture analysis.

f) Temporal Features:

- Implementation: Custom function comparing current features with previous frame's features
- Purpose: These features capture changes in facial expressions over time, potentially useful for video analysis or detecting subtle emotional changes.

Experimental Results and Evaluation Metrics

The Code conducts series of experiment with various combinations of parameters for HOG, LBP, and texture features.

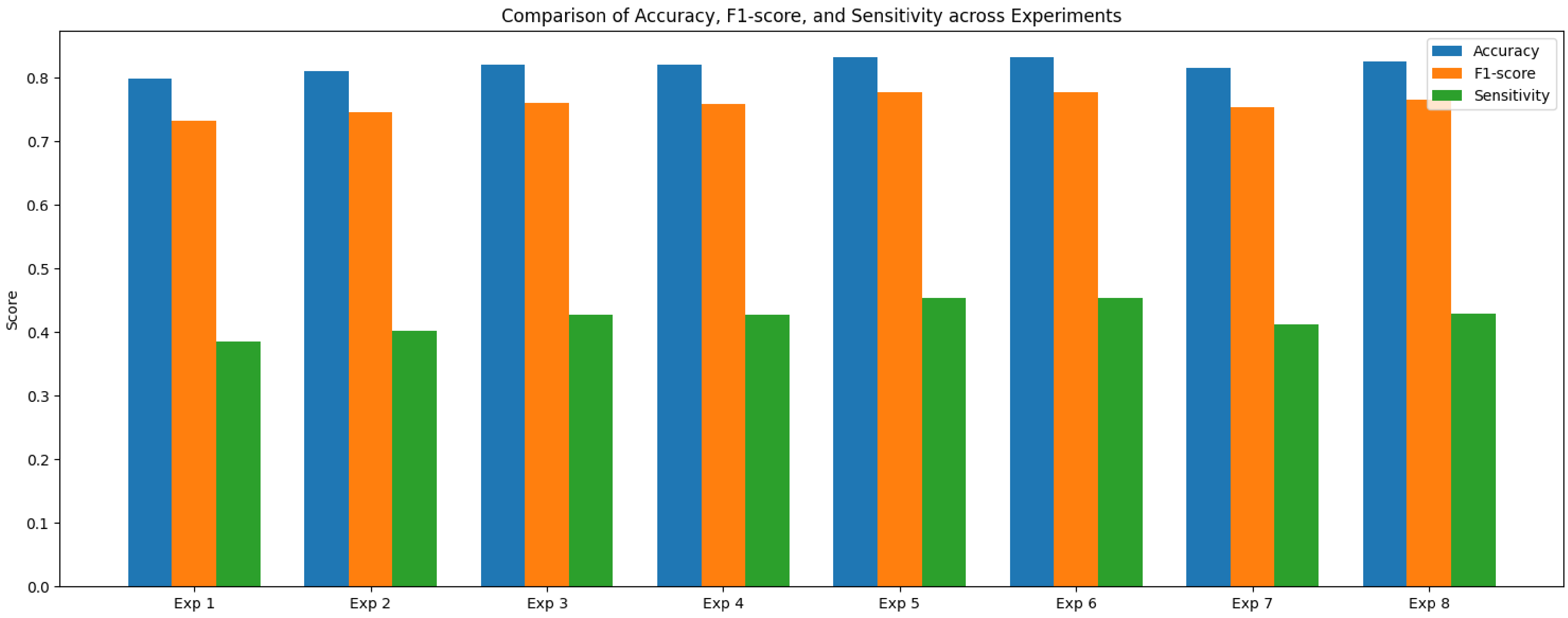
- Experiment overview:
  - 8 experiments conducted
  - Best accuracy achieved: 83.15% indicates a strong performance in facial emotion recognition
  - Overall best Sensitivity: 0.4534 suggests moderate responsiveness to changes in the input
  - Overall best F1-score: 0.7765 suggests a good balance between correctly identifying emotions (precision) and not missing instances of emotions (recall).
- Optimal Parameters:
  - HOG: Window size (64x64), larger block size and stride

    hog_params:{'win_size': (64, 64), 'block_size': (32, 32), 'block_stride': (16, 16), 'cell_size': (16, 16), 'nbins': 9}
  - LBP: Radius 1, 8 points, uniform method

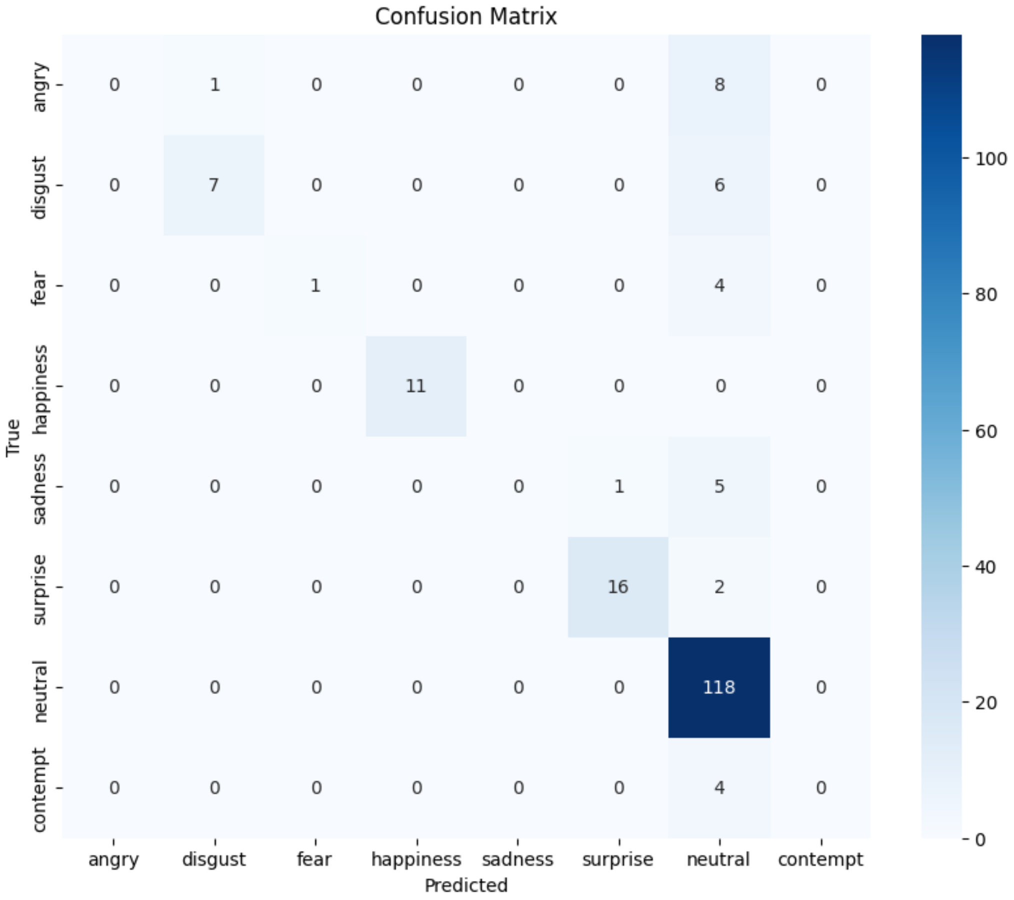
    lbp_params: {'radius': 1, 'n_points': 8, 'method': 'uniform'}
  - Texture: Distance 1, standard angles and properties

    texture_params: {'distances': [1], 'angles': [0, 0.7853981633974483, 1.5707963267948966, 2.356194490192345], 'props': ['contrast', 'dissimilarity', 'homogeneity', 'energy', 'correlation']}

Visualization of experimental results:



Confusion Matrix and Classification Report of experimental results best parameters:



	precision	recall	f1-score	support
angry	0.00	0.00	0.00	9
disgust	0.88	0.54	0.67	13
fear	1.00	0.20	0.33	5
happiness	1.00	1.00	1.00	11
sadness	0.00	0.00	0.00	6
surprise	0.94	0.89	0.91	18
neutral	0.80	1.00	0.89	118
contempt	0.00	0.00	0.00	4
accuracy			0.83	184
macro avg	0.58	0.45	0.48	184
weighted avg	0.76	0.83	0.78	184

Key Findings of experiment:

- Feature Combination Effectiveness: The multi-feature approach (HOG, LBP, facial landmarks, geometric, and texture features) achieved high accuracy, demonstrating its effectiveness in capturing diverse aspects of facial expressions.
- Global vs. Local Features: Larger HOG window size improved performance, suggesting the importance of broader facial context. Simultaneously, fine-grained LBP and texture features contributed significantly, indicating value of local detail.
- Parameter Sensitivity: Feature extraction parameters significantly impact model performance, underscoring the importance of careful tuning in facial emotion recognition tasks.
- Balance of Features: The best model struck effective balance between global structure (HOG) and local details (LBP, texture), capturing both overall facial composition and subtle expression cues.
- Potential for Improvement: While achieving good accuracy, results suggest room for further optimization through additional feature extraction techniques or fine-tuning of existing parameters.

Discussion of Strengths, Limitations, and Potential Improvements

Strengths:

- Comprehensive feature set: Code combines global features (HOG, LBP) and local characteristics (geometric, texture) to provide detailed representation of facial expressions.
- Flexibility: Implementation enables for simple experimentation with various parameter combinations, for optimization.
- Temporal consideration: Inclusion of temporal features lays groundwork for potential video-based emotion recognition.

Limitations:

- Computing complexity: Usage of various feature extraction algorithms can result in significant computing expenses, particularly for real-time applications.
- Potential overfitting: Overfitting is a danger when there are significant number of features, particularly on smaller datasets.
- Dependence on face landmark detection: The precision with which geometric features are detected has significant impact on their accuracy.

Potential Improvements:

- Feature selection: Use approaches to discover and keep only most informative characteristics, therefore lowering dimensionality and possibly enhancing generalization.
- Deep learning integration: Use of Convolutional Neural Networks (CNNs) for end-to-end learning, maybe in hybrid with handcrafted features.
- Cross-dataset validation involves testing the system on several datasets to ensure its robustness across varied populations and imaging circumstances.
- Real-time optimization: Simplify feature extraction procedure for faster processing, making code better suited to real-time applications.
- Emotion intensity: Expand code to include ability to measure the intensity of emotions in addition to classification.

Conclusion and Future Work

Conclusion:

- The built face emotion recognition system exhibits efficacy of merging different feature extraction approaches. Experiments demonstrate that careful parameter adjustment can have considerable impact on performance, with best setup obtaining 83.15% accuracy. The system provides solid foundation for emotion recognition tasks and is adaptable to a variety of settings.

Future Work:

- Develop real-time emotion recognition system for video streams, leveraging the temporal features already implemented.
- Explore transfer learning approaches using pre-trained deep learning models to potentially improve accuracy and generalization.
- Investigate the system's performance across different demographics to ensure fairness and reduce potential biases.
- Explore ensemble methods combining current feature-based approach with deep learning techniques to potentially achieve higher accuracy.
- Investigate impact of different classifiers (e.g., SVM, Neural Networks) on the performance.