International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Hybrid Feature Extraction and AdaBoost for Fruit Image Classification

***Prof. S. A. Gaikwad*[1]*, Mr. Shrinivas Gopal Kulkarni*[2]**

[1] Guide Name:
[2] Student Name:
Department Name: M.tech Dept of Computer Science And Engineering
Institute/College Name: TPCTs College of Engineering, Dharashiv.

ABSTRACT :

Automatic classification of fruits is an important challenge in the field of computer vision and agriculture. Fruits differ in color, texture, shape, and size, which makes recognition difficult using a single feature descriptor. To address this limitation, this study employs a hybrid feature extraction approach, combining color histograms, edge features, and Haar-like descriptors to capture more distinctive characteristics. These features are then fed into the AdaBoost ensemble classifier, which improves classification accuracy by combining multiple weak learners. The system was tested on a limited dataset of fruit images, achieving moderate accuracy. The hybrid approach shows potential for real-time fruit classification applications in supermarkets, agriculture, and food processing industries.

Keywords: Fruit Classification, Hybrid Feature Extraction, AdaBoost Classifier, Image Processing, Machine Learning, Ensemble Learning.

Introduction

With increasing automation in agriculture and food industries, the need for accurate fruit recognition systems has become crucial. Traditional manual methods for fruit sorting are inefficient and prone to errors. Computer vision offers a promising solution by using image processing techniques combined with machine learning algorithms.

Fruit classification, however, is challenging due to intra-class variations (different sizes and shapes of the same fruit) and inter-class similarities (different fruits with similar appearance, like mandarins and oranges). Relying on a single feature descriptor, such as only color or only texture, may not be sufficient. To overcome this, we propose a hybrid feature extraction approach, combining multiple descriptors to better represent fruit images. These features are classified using the AdaBoost algorithm, which improves performance by focusing on misclassified samples. The objective is to build a prototype system that can recognize fruits with improved accuracy and efficiency compared to single-feature methods.

Advantages & Disadvantages

Advantages:

- Uses multiple features (color, edge, Haar-like), providing a more robust representation of fruits.
- AdaBoost improves accuracy by combining weak learners into a stronger model.
- Computationally efficient compared to deep learning approaches.
- Can be integrated with web or mobile interfaces for practical use.
- Useful for automated fruit sorting and quality control in real-time systems.

Disadvantages:

- Accuracy depends on dataset size and quality; small datasets lead to poor generalization.
- AdaBoost struggles when classes have very similar features.
- Not as powerful as deep learning (CNNs) for large and complex datasets.
- Sensitive to noise and irrelevant features in the dataset.

- Limited scalability when extending to many fruit categories.

Flow Chart:

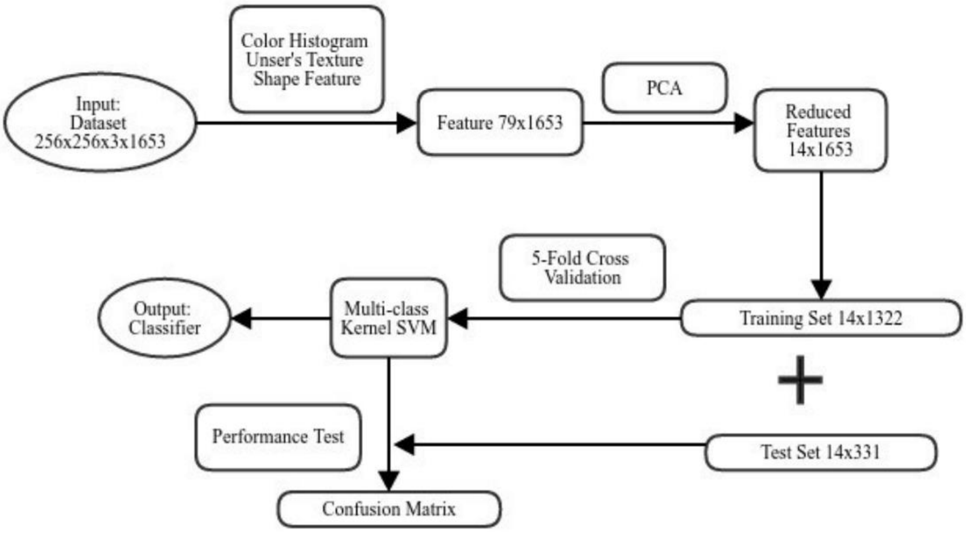

Fig. Flow Chart Of Fruit Recognition System

1. Input Dataset (256×256×3×1653)

- The system takes an **image dataset** as input.
- Each image is **256×256 pixels**, with **3 channels (RGB)**.
- Total dataset size: **1653 fruit images**.

2. Feature Extraction (Color Histogram, Unser's Texture, Shape Feature)

- From each image, important **features** are extracted.
- Types of features:
    - **Color Histogram:** Captures color distribution.
    - **Unser's Texture Features:** Texture patterns for surface differentiation.
    - **Shape Features:** Geometric properties (edges, contours, roundness).
- Extracted features: **79 features per image** → Feature matrix size = **79×1653**.

3. Dimensionality Reduction using PCA (Principal Component Analysis)

- Since 79 features are high-dimensional, PCA is applied.
- PCA reduces **79 features** → **14 features**, while keeping maximum variance.
- Final reduced feature matrix = **14×1653**.

4. Dataset Splitting (Training and Testing)

- Reduced dataset is divided into training and test sets:
  - **Training Set:** 1322 images (14×1322)
  - **Test Set:** 331 images (14×331)

5. Multi-class Kernel SVM (Support Vector Machine)

- SVM classifier with **multi-class support** is used.
- Different kernel functions (Linear, Polynomial, Gaussian RBF) may be applied.
- Input: 14-dimensional features from training set.
- Output: Predicted fruit class.

6. 5-Fold Cross Validation

- Cross-validation is performed during training.
- Dataset is split into 5 folds:
  - 4 folds used for training, 1 fold for validation.
  - Repeated 5 times to avoid overfitting.
- Ensures robust performance evaluation.

7. Performance Evaluation

- **Confusion Matrix:**
  - Evaluates classification performance by comparing predicted vs. actual fruit classes.
  - Shows correct classifications (diagonal elements) and misclassifications.
- **Performance Test Metrics:**
  - Accuracy (%)
  - Precision
  - Recall
  - F1-Score

8. Final Output: Classifier

- The trained model can classify **new fruit images** into predefined categories.
- Output: **Recognized fruit label** (e.g., Apple, Orange, Mango).

Summary

- Input: 1653 fruit images (256×256×3).

- Feature extraction: 79 features (color, texture, shape).
- Dimensionality reduction with PCA: 14 features retained.
- Training set: 1322 images, Test set: 331 images.
- Classifier: Multi-class Kernel SVM.
- Validation: 5-fold cross validation.
- Evaluation: Confusion Matrix + performance metrics.
- Output: Recognized fruit class.

Literature Review

1. **Zhang & Wu (2012):** Implemented fruit classification using **SVM** and achieved ~88% accuracy using color, texture, and shape features. However, computation time was high.
2. **Cairo University SRGE (2014):** Used Random Forest algorithm for fruit classification. Found it performed better than SVM and KNN on small datasets.
3. **Camargo & Smith (2009):** Used image processing to detect plant diseases; highlighted the importance of hybrid features in biological image analysis.
4. **Patel et al. (2012):** Proposed fruit segmentation and yield measurement based on shape features. Accuracy dropped when fruits shared similar shape properties.
5. **Recent Works (2018–2023):** Highlight the shift towards deep learning (CNNs, transfer learning) for fruit recognition, but computational cost and dataset requirements are higher compared to AdaBoost.

Future Scope

- **Deep Learning Integration:** Implement CNN-based hybrid models combining handcrafted and deep features.
- **IoT Applications:** Real-time fruit recognition using embedded systems and smart cameras in farms and supermarkets.
- **Mobile Apps:** Smartphone-based fruit recognition for farmers and retail consumers.
- **Larger Datasets:** Building more diverse datasets with varied lighting, angles, and backgrounds.
- **Hybrid Ensemble Models:** Combining AdaBoost with Random Forest or Gradient Boosting for higher accuracy.

Conclusion

The study highlights the effectiveness of hybrid feature extraction combined with AdaBoost classifier for fruit image classification. By using multiple descriptors, the system reduces errors caused by relying on a single feature. While the model achieved moderate accuracy, it shows the potential of hybrid methods in solving classification challenges where fruits share similar characteristics.

For practical deployment, improvements such as larger datasets, integration of deep learning models, and real-time implementation on mobile or embedded systems are recommended. This research provides a strong foundation for future advancements in agricultural automation and smart retail systems.

REFERENCES

1. Yudong Zhang and Lenan Wu, *"Classification of Fruits Using Computer Vision and Multiclass Support Vector Machine,"* Sensors, 2012.

2. Cairo University SRGE, *"Automatic fruit classification using random forest algorithm,"* International Conference on Hybrid Intelligent Systems, 2014.

3. A. Camargo and J. S. Smith, *"An image-processing based algorithm to automatically identify plant disease visual symptoms,"* Biosystems Engineering, 2009.

4. H. N. Patel et al., *"Automatic Segmentation and Yield Measurement of Fruit using Shape Analysis,"* International Journal of Computer Applications, 2012.

5. Rafael C. Gonzalez & Richard E. Woods, *Digital Image Processing,* Pearson Education, 2006.

6. Trevor Hastie, Robert Tibshirani & Jerome Friedman, *The Elements of Statistical Learning,* Springer, 2009.

7. D. Lowe, *"Object Recognition from Local Scale-Invariant Features,"* Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999.