Thesis submitted to the

UNIVERSITY OF MOHAMED BOUDIAF – MSILA



FACULTY OF MATHEMATICS AND COMPUTER SCIENCE
DEPARTMENT OF COMPUTER SCIENCE

In partial fulfillment of the requirements for the degree of

Master in **Computer science**

By

Bouferah, **Khireddine**

Title of the thesis

Image Recognition Using Invariant Moments

Under the supervision of

Dr. Khadidja Derdour

Composition of the jury

Dr. Hamza Loucif	University of Msila	President
Dr. Khadidja Derdour	University of Msila	Reporter
Dr. Ahlam Bouzaaroura	University of Msila	Examiner

June, 2026

Abstract

This thesis investigates the use of invariant moment descriptors for image recognition, namely Hu, Zernike, Pseudo-Zernike, Legendre, and Tchebichef moments. The study aims to evaluate their ability to represent image shapes and improve classification performance across several datasets, including MNIST, MPEG-7, FLAVIA, CIFAR-10, and others. After preprocessing, moment-based features were extracted and classified using the k-Nearest Neighbors (k-NN) classifier. The experimental results show that invariant moments are effective for image recognition, especially in shape-based tasks. Legendre moments achieved the best overall performance across most datasets, followed by Zernike and Pseudo-Zernike moments, while Hu moments generally produced lower accuracy. These findings confirm the usefulness of invariant moments in computer vision applications.

Key words: Image recognition; Invariant moments; Feature extraction; k-NN classification; Computer vision.

الملخص

تتناول هذه المذكرة استعمال معاملات العزوم الثابتة في التعرف على الصور، والمتمثلة في عزوم هو، زيرنيك، زيرنيك الزائفة، ليجاندر، وتشيبيشيف. تهدف الدراسة إلى تقييم قدرة هذه العزوم على تمثيل أشكال الصور وتحسين أداء التصنيف باستعمال عدة قواعد بيانات، منها MNIST و MPEG-7 و FLAVIA و CIFAR-10 وغيرها. بعد مرحلة المعالجة المسبقة، تم استخراج الخصائص المعتمدة على العزوم، ثم تصنيفها باستخدام خوارزمية الجيران الأقرب (k-NN). أظهرت النتائج أن العزوم الثابتة فعالة في التعرف على الصور، خاصة في المهام المعتمدة على الشكل. وقد حققت عزوم ليجاندر أفضل أداء عام في أغلب قواعد البيانات، تلتها عزوم زيرنيك وزيرنيك الزائفة، بينما سجلت عزوم هو دقة أقل. تؤكد هذه النتائج أهمية العزوم الثابتة في تطبيقات الرؤية الحاسوبية..

الكلمات المفتاحية: التعرف على الصور؛ العزوم الثابتة؛ استخراج الخصائص؛ خوارزمية الجيران الأقرب ؛ الرؤية الحاسوبية.

Dedications

" I dedicate this work to my small family, my beloved wife Souad, and my precious son Ahmed Mohamed. Your love, patience, and unwavering support have been my constant source of strength throughout this journey. This thesis is a reflection of your sacrifices and encouragement.

I also dedicate this work to my larger family, my parents and my dear siblings. Your continuous love, guidance, and belief in me have shaped the person I am today. Without your support, none of this would have been possible.

I would also like to dedicate this thesis to my friends, whose friendship, support, and encouragement have kept me going through the challenges. Your presence in my life has made this journey more meaningful, and I am truly grateful for the bond we share.

This achievement is as much yours as it is mine. Thank you for always being there for me."

Khireddine Bouferah

Acknowledgments

I would like to express my deepest gratitude to my thesis supervisor, Professor Dr. Khadidja Derdour, whose unwavering support, guidance, and continuous encouragement have been instrumental in the successful completion of this research. Her insightful feedback, dedication, and academic expertise have made this project an enriching and rewarding experience. I am sincerely grateful for her patience and commitment, which helped me navigate the challenges of this study.

I would also like to extend my heartfelt thanks to my colleagues in the Artificial Intelligence Master's program, Second Year, at the University of Msila. Their collaboration, intellectual exchange, and constant motivation have made this journey enjoyable and have broadened my horizons in many aspects of AI research.

I am deeply appreciative of the professors who have taught me and shared their invaluable knowledge throughout my academic journey. Their teachings have shaped my understanding of the field and equipped me with the necessary skills to undertake this thesis. Special thanks go to the Computer Science Department and the University of Msila for providing a stimulating academic environment that encouraged research and creativity.

Finally, I owe my profound thanks to my friends and family who have supported me throughout the ups and downs of this research journey. Their encouragement, love, and understanding during the challenging times have been my source of strength.

Once again, my heartfelt thanks to everyone who played a role in this academic and personal achievement.

Contents

List of figures ..... **vii**
List of tables ..... **viii**
Introduction **1**
Chapter 1 Image Recognition **3**
1.1. Introduction **3**
1.2. Digital Image Representation **3**
1.3. Overview of Image Recognition Systems **4**
1.3.1. Image Acquisition **4**
1.3.2. Preprocessing **4**
1.3.3. Feature Extraction Techniques **6**
A. Global Features **7**
B. Local Features **7**
C. Shape-Based Features **7**
D. Texture Features **7**
E. Color Features **7**
1.3.4. Classification **8**
A. Supervised Learning **9**
B. Unsupervised Learning **10**
1.4. Conclusion **11**
Chapter 2 Invariant Moments **13**
2.1. Introduction **13**
2.2. Geometric Image Moments **13**
2.2.1. Centroid of the Image **14**
2.2.2. Central Moments **14**
2.2.3. Normalized Central Moments **15**
2.3. Hu Invariant Moments **15**
2.4. Zernike Moments **16**
2.5. Pseudo-Zernike Moments **17**
2.6. Legendre Moments **17**
2.7. Tchebichef Moments **17**
2.8. Comparative Analysis **18**
2.9. Illustrative Experiment on Moment Invariance **20**
2.10. Conclusion **21**
Chapter 3 Application and Experimental Results **22**
3.1. Introduction **22**

3.2. Development Environment and Tools 22
3.3. General Workflow 23
3.4. Datasets Description 25
3.4.1. MNIST Dataset 26
3.4.2. MPEG-7 Dataset 26
3.4.3. FLAVIA Dataset 27
3.4.4. CIFAR-10 Dataset 28
3.4.5. Four Shapes Dataset 28
3.4.6. Vegetable Seed Image Dataset 29
3.4.7. Malaria Cell Images Dataset 29
3.4.8. GTSRB Dataset 30
3.4.9. Hand Gesture Recognition Dataset 31
3.4.10. Arabic Printed Digits Dataset 31
3.5. Preprocessing 32
3.5.1. Image Resizing 32
3.5.2. Grayscale Conversion 32
3.5.3. Image Binarization 33
3.5.4. Morphological Processing 33
3.5.5. Shape Normalization 33
3.6. Feature Extraction Methods 33
3.6.1. Hu Moments 33
3.6.2. Zernike Moments 33
3.6.3. Pseudo-Zernike Moments 34
3.6.4. Legendre Moments 34
3.6.5. Tchebichef Moments 34
3.7. Experimental Setup 34
3.7.1. Dataset Splitting 34
3.7.2. Evaluation Metrics 34
A. Accuracy 35
B. Precision 35
C. Recall 35
D. F1-score 36
3.7.3. Classification in This Work 36
A. Implementation of the k-NN Classifier 37
B. Distance Metric 37
C. Feature Standardization 37
D. Selection of k Values 37

3.8. Experimental Results 38
3.8.1. Results on MNIST 38
3.8.2. Results on MPEG-7 39
3.8.3. Results on Four Shapes 40
3.8.4. Results on FLAVIA 41
3.8.5. Results on CIFAR-10 42
3.8.6. Results on Vegetable Seeds 43
3.8.7. Results on Malaria Cell Images 44
3.8.8. Results on German Traffic Sign Recognition Benchmark 44
3.8.9. Results on Hand Gesture Recognition Dataset 45
3.8.10. Results on Arabic Printed Digits Dataset 46
3.9.Global Comparison 47
3.10. Comparison with Previous Studies 48
3.10.1. MNIST Dataset 48
A. Analysis 49
B. Discussion 50
3.10.2. Four Shapes Dataset 50
A. Analysis 50
B. Discussion 51
3.10.3. MPEG-7 Dataset 51
A. Analysis 52
B. Discussion 53
3.11. Discussion 53
3.12. Conclusion 54
Conclusion 56
References 57

List of figures

Fig.1. 1: Types of digital images..... 4
Fig.1. 2: Main preprocessing operations in an image recognition system. 5
Fig.1. 3: General image recognition pipeline showing the learning and testing phases. 8
Fig 3. 1: General workflow of the proposed image classification system……..……....24
Fig 3. 2: Graphical user interface of the developed image recognition application..... 25
Fig 3. 3: Sample images from the MNIST dataset..... 26
Fig 3. 4: Sample images from the MPEG-7 dataset..... 27
Fig 3. 5: Sample images from the FLAVIA dataset..... 27
Fig 3. 6: Sample images from the CIFAR-10 dataset. 28
Fig 3. 7: Sample images from the Four Shapes dataset. 29
Fig 3. 8: Sample images from the Vegetable Seed Image dataset. 29
Fig 3. 9: Sample images from the Malaria Cell Images dataset..... 30
Fig 3. 10: Sample images from the GTSRB dataset. 30
Fig 3. 11: Sample images from the Hand Gesture Recognition dataset..... 31
Fig 3. 12: Sample images from the Arabic Printed Digits dataset. 32

List of tables

Table 1. 1: Types of image features. 8
Table 2. 1: Comparative analysis of the studied moment descriptors……..……...19
Table 2. 2 : Hu invariant moment values for transformed horse shape samples…….20
Table 3. 1: Results on MNIST dataset..... 38
Table 3. 2: Results on MPEG-7 dataset..... 39
Table 3. 3: Results on Four Shapes dataset 40
Table 3. 4: Results on FLAVIA dataset 41
Table 3. 5: Results on CIFAR-10 dataset..... 42
Table 3. 6: Results on Vegetable Seeds dataset..... 43
Table 3. 7: Results on Malaria Cell Images..... 44
Table 3. 8: Results on German Traffic Sign Recognition Benchmark 45
Table 3. 9: Results on Hand Gesture Recognition Dataset..... 46
Table 3. 10: Results on Arabic Printed Digits Dataset 47
Table 3. 11: Overall Comparative Performance of Invariant Moment Descriptors 47
Table 3. 12: Comparison with Previous Studies on MNIST Dataset 49
Table 3. 13: Comparison with Previous Studies on Four Shapes Dataset..... 50
Table 3. 14: Comparison with Previous Studies on MPEG-7 Dataset 51

Introduction

The rapid advancement of digital technologies, combined with the widespread availability of imaging devices, has led to an exponential growth in visual data. Images are continuously generated from various sources such as surveillance systems, medical imaging devices, satellites, and multimedia platforms. This massive volume of visual information has created an urgent need for automated systems capable of analyzing and interpreting images efficiently. As a result, image recognition has emerged as a fundamental research area in computer vision and artificial intelligence[1].

Image recognition refers to the process of identifying and classifying objects, patterns, or meaningful structures within an image. It enables machines to understand visual data in a way that approximates human perception. Over the years, image recognition has found applications in many domains, including medical diagnosis, biometric systems, industrial inspection, and autonomous vehicles [2]. However, despite these advancements, designing robust recognition systems remains a challenging task due to variations in image conditions such as noise, illumination changes, and geometric transformations including translation, rotation, and scaling.

A typical image recognition system involves several stages, including image acquisition, preprocessing, segmentation, feature extraction, and classification. In this thesis, segmentation is discussed as a theoretical concept related to preprocessing, while the practical implementation focuses mainly on preprocessing, moment-based feature extraction, and classification. Among these stages, feature extraction is considered one of the most critical steps, as it directly influences the performance of the classification process [3].

Among the various feature extraction techniques, invariant moments have attracted significant attention due to their strong mathematical foundation and robustness. Invariant moments provide a numerical representation of image shapes that remains unchanged under geometric transformations such as translation, rotation, and scaling. This property makes them particularly suitable for pattern recognition tasks where objects may appear in different orientations or sizes[4].

In this work, we focus on studying and comparing different invariant moment descriptors for image recognition, namely Hu moments, Zernike moments, Pseudo-Zernike moments, Legendre moments, and Tchebichef moments. The aim is to evaluate their effectiveness in

representing image shapes and improving classification performance across datasets of varying complexity. The study considers several datasets, including MNIST, FLAVIA, Four Shapes, MPEG-7, Vegetable Seed, Malaria Cell Images, Traffic Sign Recognition (GTSRB), Hand Gesture Recognition, CIFAR-10, and the Arabic Printed Digits Dataset. These datasets were selected to cover a wide range of application domains, including handwritten digit recognition, Arabic printed digit recognition, shape analysis, plant classification, medical imaging, traffic sign recognition, gesture recognition, and natural object classification.

The contributions of this work include the implementation and comparative analysis of multiple invariant moment descriptors, the evaluation of their performance using classification metrics such as accuracy and confusion matrices, and the identification of their strengths and limitations in both simple and complex recognition scenarios.

The remainder of this thesis is organized as follows:

- Chapter 1 presents the fundamental concepts of image recognition and image processing;
- Chapter 2 introduces invariant moments and their mathematical formulations;
- Chapter 3 describes the implementation of the proposed system along with experimental results and discussion;
- Finally, a general conclusion summarizes the main findings and outlines future research directions.

Chapter 1 Image Recognition

1.1. Introduction

Image recognition is a fundamental field in computer vision that aims to enable machines to interpret and understand visual information. It involves identifying objects, patterns, or structures within digital images and assigning them to predefined categories. With the rapid growth of digital data and multimedia content, image recognition has become essential in numerous applications such as medical imaging, biometric systems, surveillance, and autonomous vehicles [5].

Unlike human vision, which relies on biological processes, image recognition systems depend on mathematical models and computational algorithms. These systems must process raw image data and transform it into meaningful information that can be used for decision-making. Designing such systems is challenging due to the complexity and variability of real-world images [6].

1.2. Digital Image Representation

A digital image can be defined as a two-dimensional function (f(x,y)), where (x) and (y) represent spatial coordinates, and the function value corresponds to the intensity at that point [7].

Images are composed of discrete units called pixels, each representing a specific intensity or color value. Depending on the type of representation, images can be classified as:

- Grayscale images, where each pixel contains a single intensity value;
- Color images, typically represented using RGB channels;
- Binary images, containing only two values (0 and 1).

Digital images can be represented in different formats depending on the application, as illustrated in Fig. 1.1.

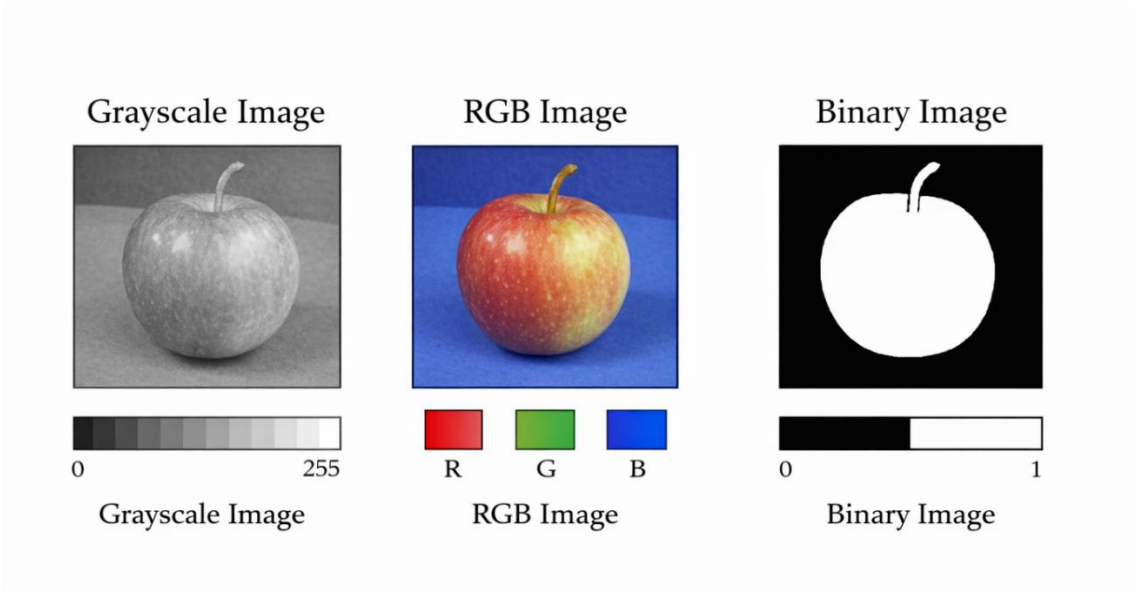


Fig.1. 1: Types of digital images

The quality and resolution of an image directly influence the performance of recognition systems. Higher resolution images provide more detailed information but increase computational complexity.

1.3. Overview of Image Recognition Systems

The general process of an image recognition system consists of several sequential stages, including image acquisition, preprocessing, feature extraction, classification, and decision-making. In the general image recognition pipeline, segmentation may be included as a preprocessing-related operation because it helps isolate the object of interest before feature extraction. However, in this thesis, segmentation is presented from a theoretical perspective and is not treated as a separate experimental stage.

The pipeline typically includes the following steps:

1.3.1. Image Acquisition

Images are captured using sensors such as cameras or scanners. The quality of acquisition directly affects system performance.

1.3.2. Preprocessing

Preprocessing is a fundamental stage in image recognition systems because it prepares the acquired image for subsequent analysis, feature extraction, and classification. Its main objective is to improve image quality, reduce irrelevant distortions, and emphasize the visual information

that is most useful for recognition. In digital image processing, preprocessing is considered part of the low-level operations applied before higher-level interpretation and decision-making.

The principal preprocessing operations commonly used in image recognition systems are illustrated in Fig. 1.2.

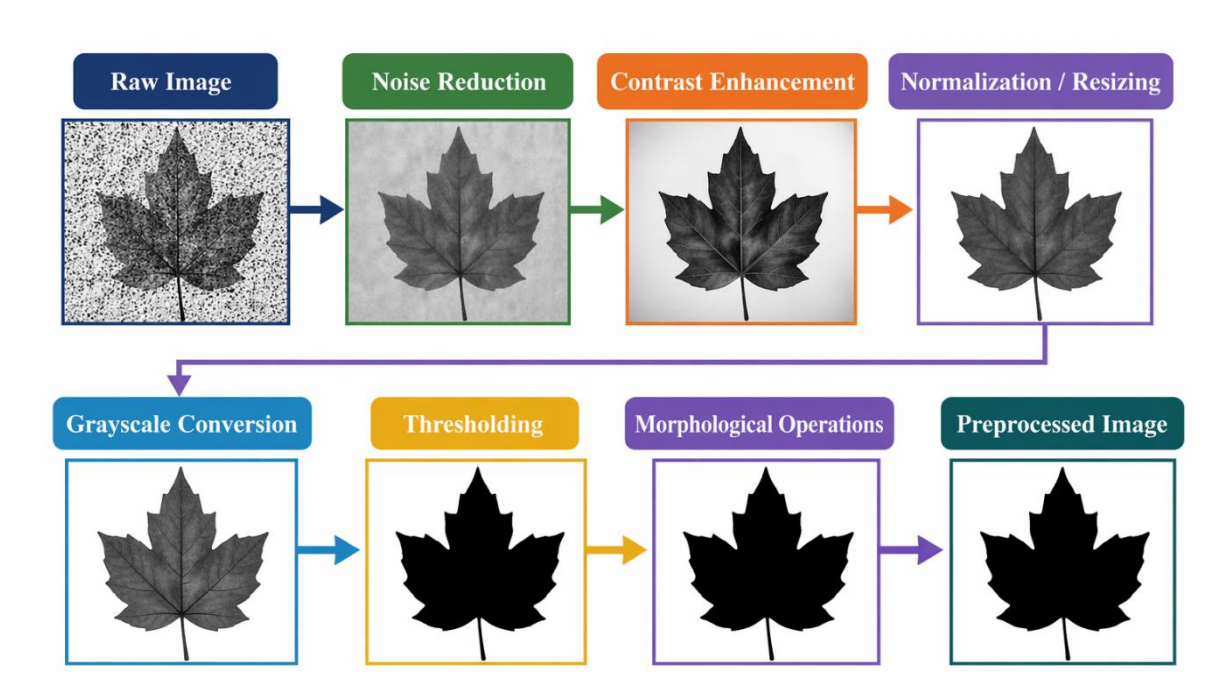


Fig.1. 2: Main preprocessing operations in an image recognition system.

As shown in Fig. 1.2, preprocessing transforms the raw input image into a cleaner and more standardized representation that is more suitable for feature extraction and classification.

Noise reduction is one of the most common preprocessing operations. Its purpose is to suppress unwanted fluctuations introduced during image acquisition, transmission, or sensor limitations. This is typically achieved using filtering techniques such as smoothing, Gaussian filtering, or median filtering. These methods reduce random disturbances while preserving the important structural information of the image.

Another important preprocessing operation is contrast enhancement. Its objective is to increase the visual distinction between the object of interest and the background, making the relevant structures easier to detect and analyze. Enhancement techniques may include intensity transformations, histogram-based methods, and spatial filtering. These operations are particularly useful when the original image is poorly illuminated or has low contrast.

Normalization is also a key step in preprocessing. It is used to standardize the image representation by reducing variations related to scale, illumination, or intensity range. In many

image recognition systems, resizing is first applied so that all input images have the same dimensions, which facilitates subsequent processing and ensures the comparability of extracted features. In addition, grayscale conversion is often used when color information is not essential, since it reduces computational complexity while preserving the structural content of the image.

For shape-based recognition, thresholding and morphological operations are particularly important. Thresholding transforms a grayscale image into a binary image by separating the foreground from the background according to intensity values. Morphological operations such as erosion, dilation, opening, and closing are then used to remove small artifacts, fill gaps, smooth boundaries, and preserve the overall geometric structure of the object. These operations are especially useful when recognition depends mainly on contour and shape information.

From a theoretical point of view, segmentation can be considered an important preprocessing-related concept, especially in shape-based image recognition. Segmentation divides the image into meaningful regions and isolates the object of interest from the background. This step allows the system to focus on the relevant object while reducing the influence of irrelevant image regions. In the case of invariant moment descriptors, segmentation is particularly important because these descriptors depend strongly on the geometric structure of the object. Therefore, a well-segmented image improves the quality of shape representation and contributes to more reliable feature extraction and classification.

Therefore, preprocessing is not merely an auxiliary step, but a decisive stage that directly influences the robustness and accuracy of the recognition system. A suitable preprocessing strategy improves the reliability of the extracted features and reduces the negative effects of noise, contrast variation, and background interference. This is especially important in systems based on invariant moments, where the quality of shape representation strongly depends on the quality of the preprocessed image [8].

1.3.3. Feature Extraction Techniques

Feature extraction is a fundamental stage in image recognition because it converts image data into a set of meaningful numerical descriptors that can be used by classification algorithms. Instead of processing all pixels directly, feature extraction aims to represent the image through compact and discriminative characteristics such as shape, texture, color, edges, or local structures. The quality of the extracted features strongly influences the accuracy and robustness of the recognition system [9].

Image features can be classified into several categories depending on the type of visual information they describe.

A. Global Features

Global features describe the image as a whole by capturing its overall characteristics. These features are computed from the entire image and provide a compact representation. Examples include shape descriptors and moment-based features such as Hu, Zernike, and Legendre moments.

B. Local Features

Local features focus on specific regions or key points within the image rather than the entire image. These features are particularly useful in complex scenes where objects may vary in position or orientation. Examples include edge-based features, corner detection methods, and keypoint descriptors such as SIFT and SURF.

C. Shape-Based Features

Shape-based features describe the geometric structure of objects, including contours and regions. These features are essential in applications where object shape is the primary discriminative factor. Moment-based descriptors belong to this category and provide invariance to translation, rotation, and scaling.

D. Texture Features

Texture features represent the spatial arrangement of pixel intensities and are useful for describing surface patterns. Common techniques include Gray-Level Co-occurrence Matrix (GLCM), Local Binary Patterns (LBP), and Gabor filters.

E. Color Features

Color features describe the distribution of colors in an image and are useful in applications where color information is important. These features are typically extracted using color histograms or color space transformations such as RGB and HSV.

In this work, the focus is placed on global shape-based features, specifically invariant moment descriptors. These features provide a compact and robust representation of object shape and are well suited for image recognition tasks where geometric properties play a dominant role.

A summary of the main types of image features is presented in Table 1.1.

Table 1. 1: Types of image features.

Feature Type	Description	Examples
Global Features	Describe the whole image	Moments, statistical features
Local Features	Describe specific regions	SIFT, SURF, edges, corners
Shape Features	Represent object geometry	Hu, Zernike, Legendre moments
Texture Features	Describe surface patterns	GLCM, LBP, Gabor filters
Color Features	Represent color distribution	Color histograms, RGB, HSV

1.3.4. Classification

Classification is the decision-making stage of an image recognition system. After preprocessing and feature extraction, each image is represented by a numerical feature vector that must be assigned to a specific class or grouped according to similarity. The classification process depends on the nature of the available data and on whether class labels are provided or not.

This stage can be better understood through the two main phases of an image recognition system: the learning phase and the testing phase, as illustrated in Fig. 1.3.

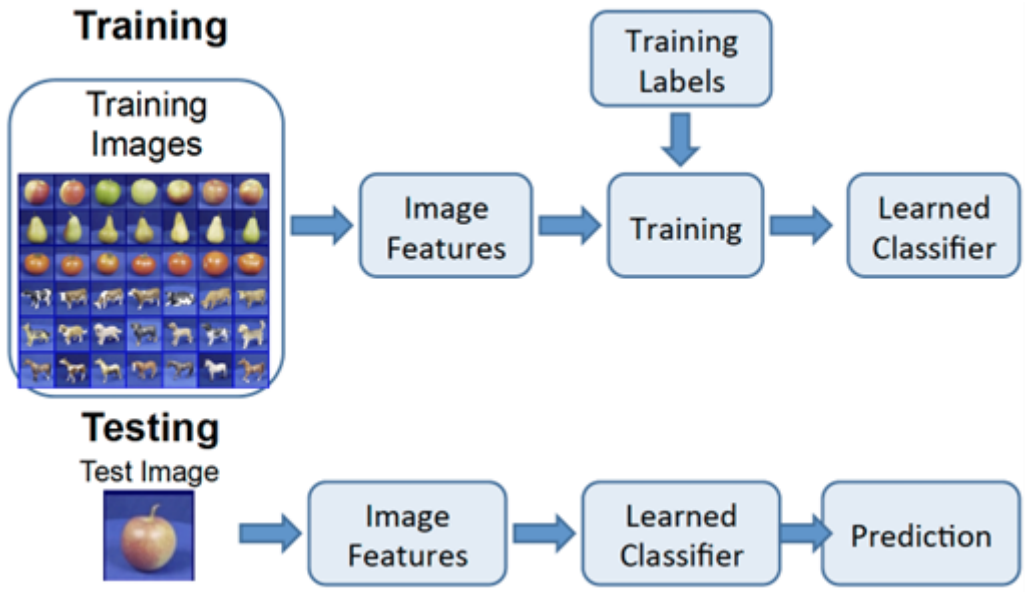


Fig.1. 3: General image recognition pipeline showing the learning and testing phases[10].

During the learning phase, also called the training phase, the system uses a set of training images with known class labels. Each training image is processed through the previous stages, including preprocessing and feature extraction. The extracted feature vectors are then associated with their corresponding labels, allowing the classifier to learn the relationship between image features and image classes. This phase enables the system to build a decision rule or a learned model that can be used later for recognition.

During the testing phase, a new unseen image is introduced into the system. This image undergoes the same preprocessing and feature extraction operations applied during the learning phase. The resulting feature vector is then submitted to the learned classifier, which predicts the most appropriate class according to the information learned during training. Therefore, the testing phase evaluates the ability of the recognition system to correctly classify new images that were not used during learning.

In image recognition, classification approaches are generally divided into two main categories: supervised learning and unsupervised learning. Supervised learning is used when labeled data are available, whereas unsupervised learning is used when the objective is to discover hidden structures or natural groups within unlabeled data.

A. Supervised Learning

Supervised learning is the most widely used approach in image recognition systems. In this paradigm, the model is trained using a labeled dataset, where each input feature vector is associated with a known class label. The objective is to learn a mapping function that can correctly assign labels to new unseen data.

In supervised image classification, the training data provide examples of the relationship between extracted features and their corresponding classes. Once this relationship is learned, the classifier can predict the class of a new image based on its feature representation.

Several machine learning algorithms are commonly used in supervised classification [11].

– k-Nearest Neighbors (k-NN)

The k-Nearest Neighbors algorithm is a simple and effective supervised classification method. It classifies a new sample according to the majority class of its k nearest neighbors in the feature space. The distance between feature vectors is usually computed using a similarity measure such as Euclidean distance.

Unlike many learning algorithms, k-NN does not require an explicit training phase. Instead, it stores the training samples and compares each new test sample with them during classification. This characteristic makes k-NN particularly useful for evaluating the quality of extracted features. If the features are discriminative, samples belonging to the same class will be close to each other in the feature space [12].

- **Support Vector Machines (SVM)**

Support Vector Machines are supervised learning models that aim to find an optimal decision boundary between classes. This boundary, called a hyperplane, is selected in such a way that the margin between different classes is maximized.

SVM is highly effective in high-dimensional feature spaces and has been widely used in pattern recognition and image classification tasks. It is especially useful when the number of features is large and when clear separation between classes is required [13].

- **Decision Trees**

Decision Trees classify data by learning a set of decision rules based on feature values. The model is structured as a tree, where each internal node represents a test on a feature, each branch represents the result of that test, and each leaf node represents a class label.

Decision Trees are easy to interpret and can handle both numerical and categorical data. However, they may be sensitive to overfitting if the tree becomes too complex [14].

- **Artificial Neural Networks (ANN)**

Artificial Neural Networks are supervised learning models inspired by the structure of the human brain. They consist of interconnected layers of artificial neurons capable of learning complex nonlinear relationships between input features and output classes.

ANNs, particularly deep neural networks, are widely used in modern image recognition systems because of their ability to learn hierarchical feature representations directly from image data. However, they generally require large datasets, high computational resources, and longer training times [15].

B. Unsupervised Learning

Unsupervised learning is used when labeled data are not available. In this approach, the objective is not to assign predefined class labels, but rather to discover hidden structures, similarities, or patterns within the data.

In image analysis, unsupervised learning methods are commonly used for exploratory analysis, data grouping, clustering, and pattern discovery. These methods can reveal natural groupings among images based on their extracted features.

The most common unsupervised methods include k-means clustering, hierarchical clustering, and Gaussian Mixture Models.

- **k-means Clustering**

k-means clustering partitions the data into k clusters by minimizing the distance between samples and their corresponding cluster centers. Each sample is assigned to the nearest cluster center, and the centers are updated iteratively until convergence.

This method is simple and efficient, but it requires the number of clusters to be specified in advance.

- **Hierarchical Clustering**

Hierarchical clustering builds a hierarchy of clusters. It can be performed using an agglomerative approach, where small clusters are progressively merged, or a divisive approach, where large clusters are progressively split.

This method is useful when the structure of the data must be analyzed at different levels of granularity.

- **Gaussian Mixture Models (GMM)**

Gaussian Mixture Models are probabilistic models that assume the data are generated from a mixture of several Gaussian distributions. Each Gaussian component represents a cluster, and each sample is assigned to clusters according to probability values.

GMMs are more flexible than k-means because they can model clusters with different shapes and distributions.

1.4. Conclusion

In this chapter, we presented the fundamental steps of image recognition, including digital image representation, preprocessing, feature extraction techniques, and classification methods. We also discussed the main challenges associated with this field.

Feature extraction remains a critical component in image recognition systems, and the choice of descriptors significantly influences performance. Invariant moments provide a powerful solution for shape representation, which will be explored in detail in the next chapter.

Chapter 2 Invariant Moments

2.1. Introduction

Feature extraction is a fundamental stage in image recognition systems, as it transforms raw image data into meaningful numerical descriptors that can be used for classification. Among the various feature extraction techniques, invariant moments have gained significant importance due to their ability to represent object shapes while remaining invariant to geometric transformations such as translation, rotation, and scaling [16].

In practical applications, objects rarely appear in fixed positions or orientations. Therefore, designing descriptors that are insensitive to such variations is essential. Invariant moments provide a mathematical framework that ensures stable and consistent representation of shapes regardless of their spatial transformations.

2.2. Geometric Image Moments

The geometric moment of order (p + q) for a two-dimensional image function f(x,y) is defined as:

$$m_{pq} = \sum_x \sum_y x^p y^q f(x, y) \tag{2.1}$$

Where:

- m_{pq}: moment of order (p + q).
- x, y: spatial coordinates of the pixel.
- p, q: non-negative integers representing the order of the moment.
- f(x,y): intensity value of the pixel at position (x,y).

This equation represents a weighted summation of pixel intensities. Each pixel contributes to the total moment according to its position and intensity. Lower-order moments describe basic properties of the image, while higher-order moments capture more complex shape characteristics.

The zeroth-order moment is given by:

$$m_{00} = \sum_x \sum_y f(x, y) \tag{2.2}$$

This moment represents the total intensity of the image, and in binary images, it corresponds to the area (number of object pixels).

2.2.1. Centroid of the Image

The centroid (center of mass) of the image is computed using first-order moments:

$$\bar{x} = \frac{m_{10}}{m_{00}} \tag{2.3}$$

$$\bar{y} = \frac{m_{01}}{m_{00}} \tag{2.4}$$

Where:

- $\bar{x}, \bar{y}$: coordinates of the centroid;
- m_{10}, m_{01}: first-order moments;
- m_{00}: total area.

The centroid represents the average position of all pixels in the image. It serves as a reference point for achieving translation invariance.

2.2.2. Central Moments

Geometric moments are sensitive to translation. To eliminate the effect of object displacement, central moments are defined as:

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y) \tag{2.5}$$

Where:

- μ_{pq}: central moment of order (p + q);
- $\bar{x}$,$\bar{y}$: centroid coordinates.

In this equation, pixel coordinates are shifted relative to the centroid. This ensures that the computed moments remain unchanged when the object moves within the image. Therefore, central moments are invariant to translation [17].

2.2.3. Normalized Central Moments

Although central moments remove translation effects, they are still sensitive to scaling. To achieve scale invariance, normalized central moments are defined as:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{oo}^{\gamma}} \tag{2.6}$$

Where:

$$\gamma = \frac{p+q}{2} + 1 \tag{2.7}$$

Where:

- η_{pq}: normalized central moment;
- μ_{pq}: central moment;
- μ_{00}: area of the object;
- γ: normalization factor.

This normalization removes the dependency on object size. As a result, normalized moments remain invariant when the object is resized [18].

2.3. Hu Invariant Moments

Hu introduced seven invariant moments derived from normalized central moments. These moments are invariant to translation, rotation, and scaling [19].

The first Hu moment is defined as:

$$\phi_1 = \eta_{20} + \eta_{02} \tag{2.8}$$

This moment measures the overall spread of the object around its centroid.

The second moment is:

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \tag{2.9}$$

This moment captures the difference between horizontal and vertical distributions and is sensitive to object orientation.

The third moment is:

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \tag{2.10}$$

The fourth moment is:

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \quad (2.11)$$

These higher-order moments capture asymmetry and shape irregularities.

The remaining Hu moments are defined as follows:

$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (2.12)$$

$$\phi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \quad (2.13)$$

$$\phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (2.14)$$

These higher-order moments capture complex geometric relationships between different regions of the shape, but they are more sensitive to noise.

2.4. Zernike Moments

Zernike moments are orthogonal moments defined over the unit circle:

$$Z_{nm} = \frac{n+1}{\pi} \sum_x \sum_y f(x,y) V_{nm}^*(x,y) \quad (2.15)$$

Where:

- Z_{nm}: Zernike moment;
- n: order of the polynomial;
- m: repetition factor;
- V_{nm}: Zernike polynomial;
- *: complex conjugate.

The Zernike polynomial is defined as:

$$V_{nm}(r,\theta) = R_{nm}(r) e^{jm\theta} \quad (2.16)$$

Where:

- r: radial distance;
- θ: angle;

- R_{nm}: radial polynomial.

Zernike moments are inherently rotation invariant due to their polar representation. The magnitude of Zernike moments is often used as a feature descriptor [20].

2.5. Pseudo-Zernike Moments

Pseudo-Zernike moments extend Zernike moments and provide improved robustness:

$$P_{nm} = \frac{n+1}{\pi} \sum_x \sum_y f(x,y) R_{nm}(r) e^{-jm\theta} \quad (2.17)$$

They provide:

- better noise resistance;
- improved numerical stability;
- higher feature representation capability [21].

2.6. Legendre Moments

Legendre moments are defined using orthogonal polynomials:

$$L_{pq} = \sum_x \sum_y P_p(x) P_q(y) f(x,y) \quad (2.18)$$

Where:

- $P_p(x)$: Legendre polynomial of order p

Legendre moments are computationally efficient and operate directly in the spatial domain [22].

2.7. Tchebichef Moments

Tchebichef moments are discrete orthogonal moments defined as:

$$T_{pq} = \sum_x \sum_y t_p(x) t_q(y) f(x,y) \quad (2.19)$$

Where:

- $t_p(x)$: discrete Tchebichef polynomial.

These moments are particularly suitable for digital images because they do not require approximation [23].

2.8. Comparative Analysis

The moment descriptors presented in this chapter exhibit different properties in terms of invariance, robustness, computational complexity, and suitability for shape representation. A comparative analysis is therefore necessary in order to highlight their strengths and limitations and to better understand their role in image recognition tasks.

Geometric moments constitute the basis of moment theory and provide useful information about the spatial distribution of image intensities. However, they are not inherently invariant to geometric transformations such as translation, scaling, and rotation. For this reason, they are generally considered as fundamental descriptors from which more advanced invariant moments are derived.

Hu invariant moments are derived from normalized central moments and provide invariance to translation, scaling, and rotation. Their main advantage lies in their simplicity and low computational cost, which makes them attractive for basic shape recognition tasks. However, Hu moments are often sensitive to noise and may offer limited discrimination power when dealing with complex shapes.

Zernike moments are orthogonal moments defined on the unit circle. Their orthogonality reduces redundancy among features and allows a compact and efficient representation of shape information. In addition, Zernike moments are known for their robustness to rotation and their strong discriminative capability in pattern recognition tasks.

Pseudo-Zernike moments are an extension of Zernike moments. They provide a larger set of descriptors and generally offer better robustness to noise. Because of their higher descriptive capacity, they are often more effective in representing fine shape variations and subtle differences between image classes.

Legendre moments are also orthogonal and provide an efficient representation of image content. They are characterized by good approximation properties and relatively moderate computational complexity. Although they are not inherently rotation invariant, they remain effective for many recognition tasks when combined with suitable preprocessing or normalization steps.

Tchebichef moments belong to the class of discrete orthogonal moments and are particularly suitable for digital image analysis. Since they are defined directly in the

discrete domain, they avoid approximation errors associated with continuous formulations and are well adapted to digital images. They also offer good numerical stability and efficient computation.

A summary of the main characteristics of the studied moment descriptors is presented in Table 2.1.

Table 2. 1: Comparative analysis of the studied moment descriptors.

Descriptor	**Main Properties**	**Advantages**	**Limitations**
Geometric Moments	Basic image moments describing spatial intensity distribution	Simple mathematical foundation	Not invariant to translation, scaling, or rotation
Hu Moments	Derived from normalized central moments; invariant to translation, scaling, and rotation	Simple, low computational cost, widely used	Sensitive to noise, limited discrimination power
Zernike Moments	Orthogonal moments defined on the unit circle	Low redundancy, strong discrimination, rotation robustness	Higher computational complexity
Pseudo-Zernike Moments	Extension of Zernike moments with more descriptors	Better robustness to noise, higher descriptive capability	Higher computational cost
Legendre Moments	Orthogonal polynomial-based moments	Efficient representation, good approximation properties	Not inherently rotation invariant
Tchebichef Moments	Discrete orthogonal moments	Suitable for digital images, numerically stable, efficient	May require additional normalization for invariance

As shown in Table 2.1, Hu moments are computationally simple but less robust to noise, whereas orthogonal moments such as Zernike, Pseudo-Zernike, Legendre, and Tchebichef provide a more compact and less redundant representation. Among these descriptors, Pseudo-Zernike moments generally offer stronger robustness, while Tchebichef moments are particularly suitable for discrete image representation.

Overall, the choice of a moment descriptor depends on the requirements of the application, including invariance properties, robustness, computational efficiency, and compatibility with digital image representation. In this work, these descriptors are experimentally evaluated in order to determine their effectiveness in image recognition tasks across multiple datasets.

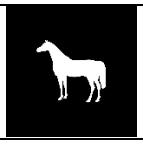
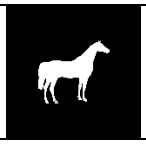
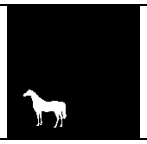
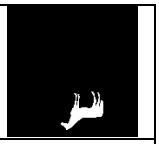
2.9. Illustrative Experiment on Moment Invariance

In order to provide a clearer understanding of the concept of invariance, an illustrative experiment was conducted using different transformed versions of the same horse shape. The objective of this experiment is not to perform classification, but to show how invariant moment descriptors can preserve relatively stable numerical values when the same object is subjected to geometric transformations.

Six images representing the same horse shape were prepared in different forms, including changes in position, orientation, and size. Each image was converted into a binary representation and then processed using the same Hu moment extraction procedure. Hu invariant moments were selected for this illustrative example because they are simple, compact, and widely used to demonstrate invariance to translation, rotation, and scaling.

The obtained Hu moment values are presented in Table 2.2. The images placed in the first row of the table correspond to the six transformed versions of the same horse shape. Since the values were computed using a logarithmic transformation, the comparison focuses mainly on the relative stability of the values rather than their absolute magnitude.

Table 2. 2 : Hu invariant moment values for transformed horse shape samples.

Hu Moment						
$\phi 1$	-0.518151	-0.529910	-0.538851	-0.517687	-0.566705	-0.556407
$\phi 2$	-1.332429	-1.356094	-1.405409	-1.323946	-1.471923	-1.401835
$\phi 3$	-2.463642	-2.640960	-2.479653	-2.502671	-2.725435	-2.732681
$\phi 4$	-3.245354	-3.379810	-3.488006	-3.312512	-3.441836	-3.558261
$\phi 5$	-6.144166	-6.402861	-6.608804	-6.272688	-6.551166	-6.935951
$\phi 6$	-4.292047	-4.321531	-5.354048	-4.446118	-4.473814	-5.683172
$\phi 7$	6.466742	7.013556	6.636799	-6.553812	7.001651	6.795019

The values shown in Table 2.2 indicate that the Hu moment descriptors remain relatively close for the transformed versions of the same horse shape. In particular, the

first Hu moments, especially $\phi 1$ and $\phi 2$, show limited variation across the six samples. This confirms their ability to preserve the global shape information of the object under changes in position, orientation, and size.

Some variations can still be observed in the higher-order moments, especially $\phi 5$, $\phi 6$, and $\phi 7$. These variations are expected because higher-order moments are more sensitive to small changes in contour shape, binarization, interpolation, and preprocessing operations. In addition, the sign of some Hu moments may change when the contour structure is slightly affected by rotation or scaling. Therefore, moment invariance in practical digital images should be understood as relative stability rather than exact numerical equality.

Overall, this illustrative experiment supports the theoretical concept of invariant moments. It shows that the same object, even when represented under different geometric conditions, can produce moment values that remain relatively stable. This explains why invariant moments are useful for shape-based image recognition tasks, especially when objects may appear with different positions, orientations, or sizes.

2.10. Conclusion

Invariant moments provide a powerful mathematical representation of shapes in image recognition systems. Their invariance properties make them essential for robust classification under varying conditions. These descriptors will be applied and evaluated in the next chapter.

Chapter 3 Application and Experimental Results

3.1. Introduction

This chapter presents the practical implementation of the proposed system for shape recognition and classification using moment-based descriptors. The work involves the development of a Python-based application for feature extraction, classification, and evaluation across multiple datasets.

The chapter details the system architecture, datasets, tools used, feature extraction methods, experimental setup, and obtained results.

3.2. Development Environment and Tools

The implementation of the proposed system was carried out using **Python 3.11.0** within the **PyCharm 2025.2.4** development environment. Several image datasets of varying complexity were used to evaluate the robustness and generalization capability of the proposed approach.

The implementation of the system was conducted using the following tools:

- **Programming Language:** Python 3.11.0;
- **IDE:** PyCharm 2025.2.4;
- **Libraries Used:**
  - ✓ OpenCV (image processing);
  - ✓ NumPy (numerical computation);
  - ✓ Scikit-learn (classification and evaluation);
  - ✓ Matplotlib (visualization);
  - ✓ PIL (image handling).

The experiments were performed on a Dell Latitude 5400 laptop with the following hardware and software specifications:

- **Operating System**: Windows 11 Pro 64-bit;
- **Processor:** Intel(R) Core(TM) i5-8365U CPU @ 1.60 GHz;
- **Memory (RAM):** 16 GB;
- **Graphics Card:** Intel(R) UHD Graphics;

- **Approximate Total Graphics Memory:** 8227 MB;
- **Dedicated Video Memory (VRAM):** 128 MB;
- **Shared Graphics Memory:** 8099 MB;
- **Display Resolution:** 1920 × 1080, 60 Hz;
- **DirectX Version:** DirectX 12.

These specifications provided a suitable development and execution environment for implementing the application, running the experiments, and obtaining the classification results presented in this chapter.

3.3. General Workflow

The proposed system follows a structured pipeline consisting of the following main stages:

1. Image Input;
2. Preprocessing;
3. Feature Extraction using Moment Descriptors;
4. Classification;
5. Evaluation.

This workflow is illustrated in Fig. 3.1, which presents the general architecture of the proposed image recognition system.

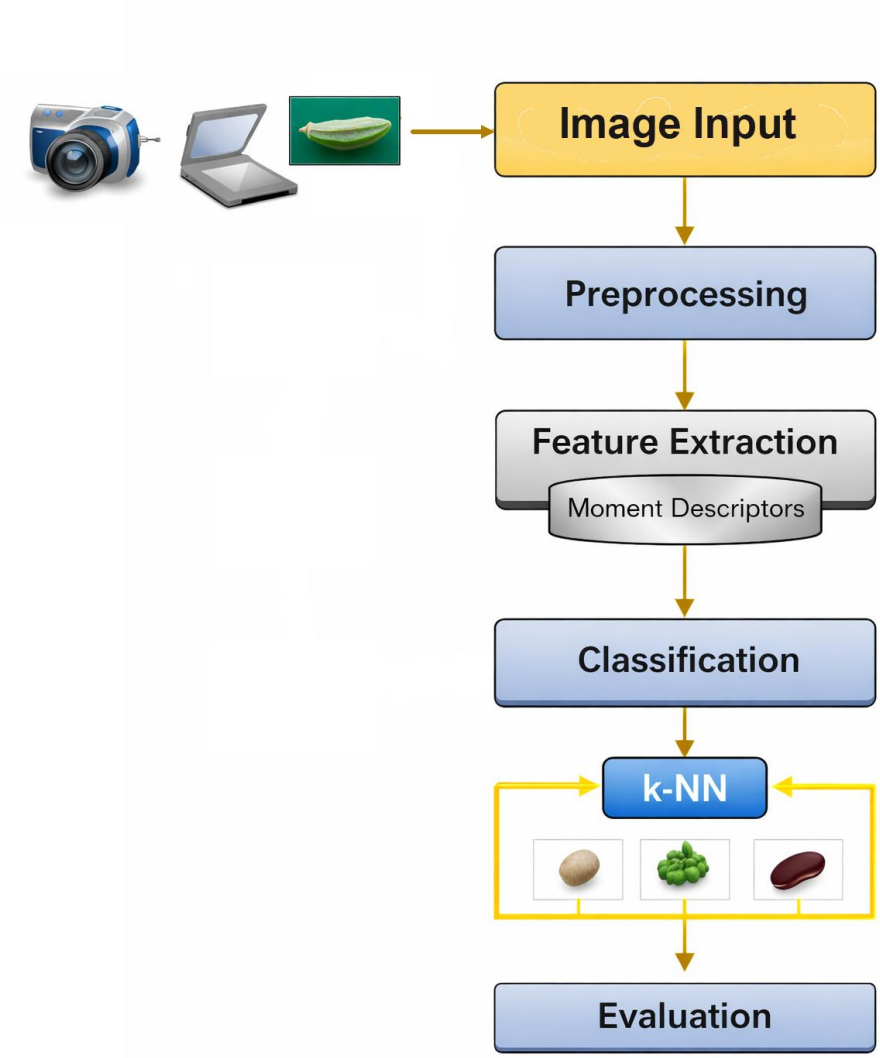


Fig 3. 1: General workflow of the proposed image classification system.

To implement the proposed workflow, a graphical application was developed for dataset handling, preprocessing, feature extraction, classification, and result visualization. The graphical user interface of the developed application is shown in Fig. 3.2.

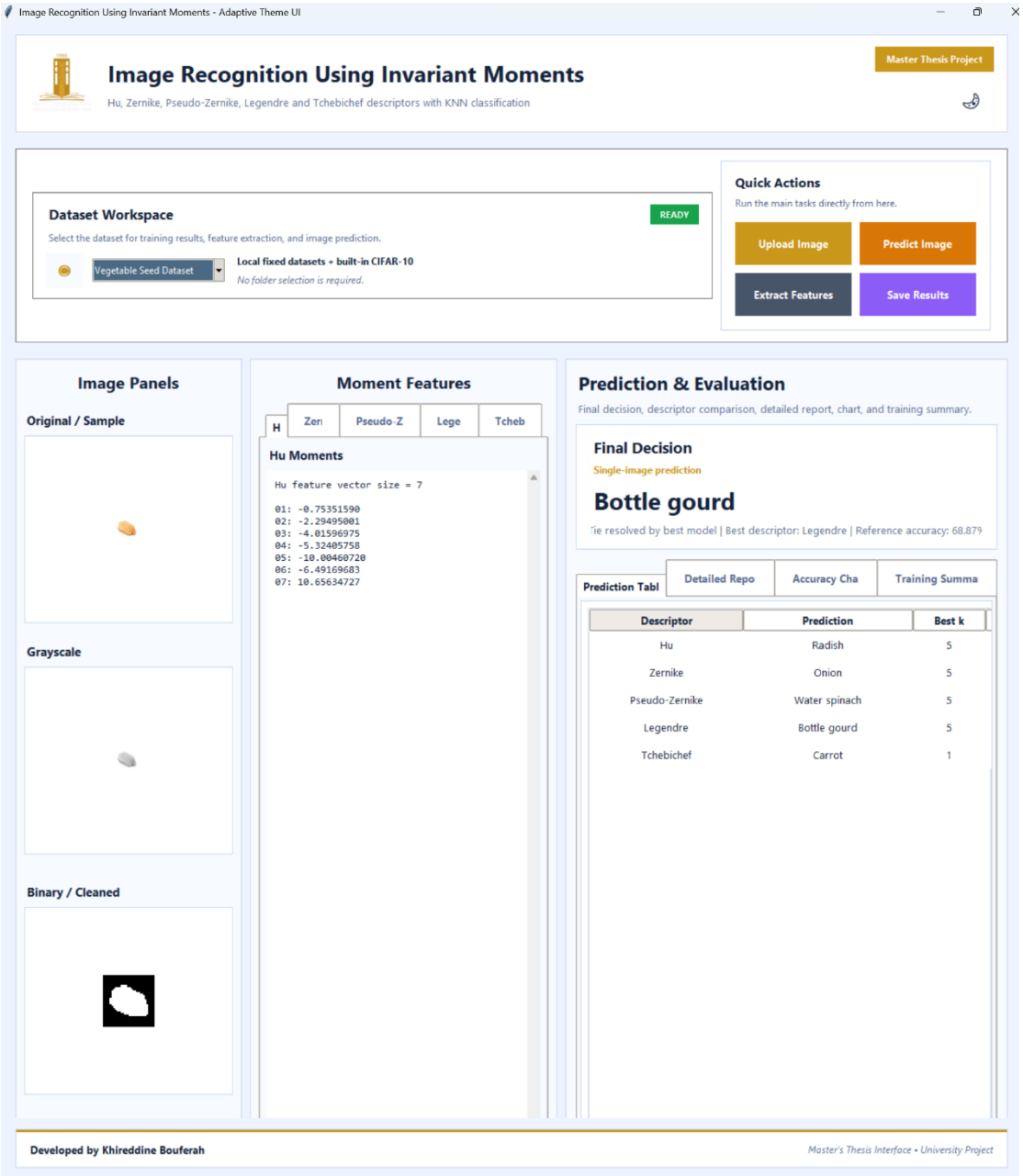


Fig 3. 2: Graphical user interface of the developed image recognition application.

3.4. Datasets Description

To ensure a comprehensive evaluation of the proposed recognition system, ten different datasets were used in this work. These datasets were selected to cover a wide range of image recognition scenarios, including handwritten digits, printed digits, shape-based silhouettes, plant leaves, natural object images, biological seed images, medical cell images, traffic sign images, and hand gesture images. This diversity makes it possible to evaluate the robustness, discriminative capability, and generalization performance of the studied moment descriptors under different levels of complexity.

3.4.1. MNIST Dataset

The MNIST dataset is one of the most widely used benchmarks in image recognition and handwritten digit classification. It contains grayscale images of handwritten digits from 0 to 9, with a total of 70,000 images divided into 60,000 training images and 10,000 test images. Each image has a size of 28 × 28 pixels. In this work, MNIST was used to evaluate the behavior of the studied descriptors on simple and well-structured shapes [24].

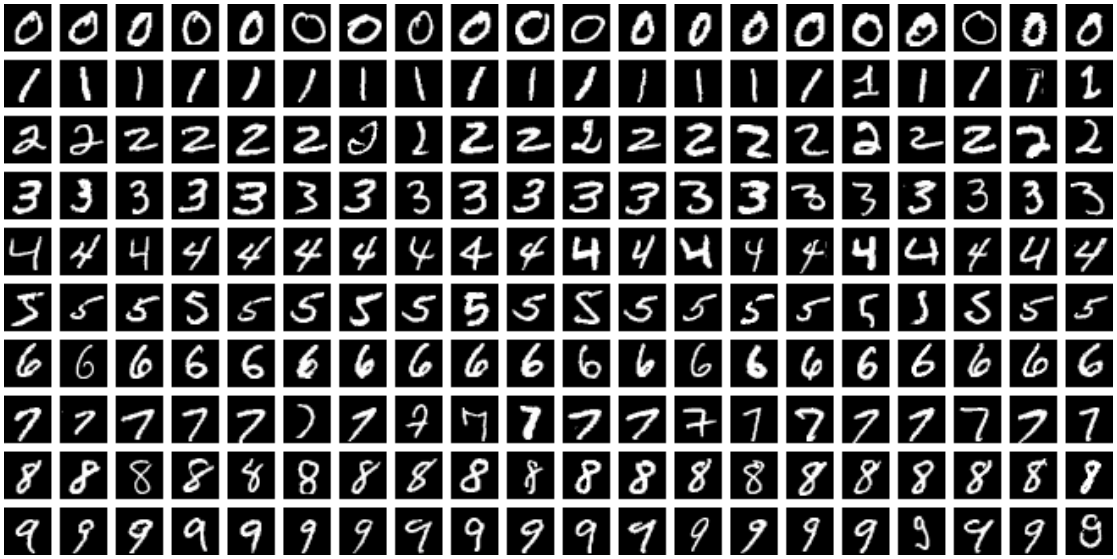


Fig 3. 3: Sample images from the MNIST dataset.

3.4.2. MPEG-7 Dataset

The MPEG-7 dataset contains 1,400 binary silhouette images distributed over 70 classes, with 20 images per class. It includes challenging geometric variations such as rotation, scaling, and deformation. This dataset was used to evaluate the effectiveness of the studied descriptors in shape recognition tasks [25].

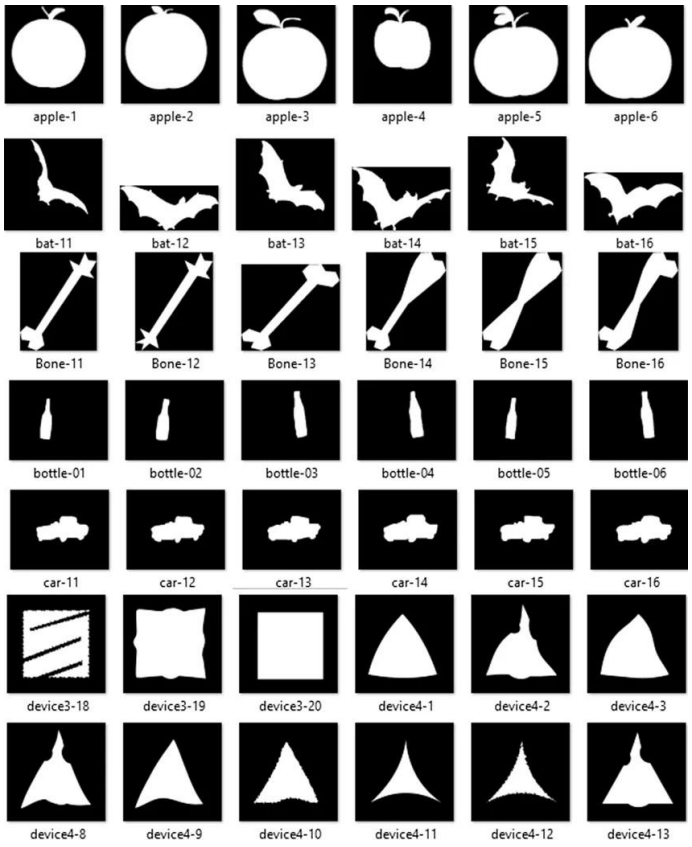


Fig 3. 4: Sample images from the MPEG-7 dataset.

3.4.3. FLAVIA Dataset

The FLAVIA dataset contains 1,907 leaf images belonging to 32 plant species. It was used to test the proposed system on natural biological shapes and evaluate descriptor performance in plant classification tasks [26].

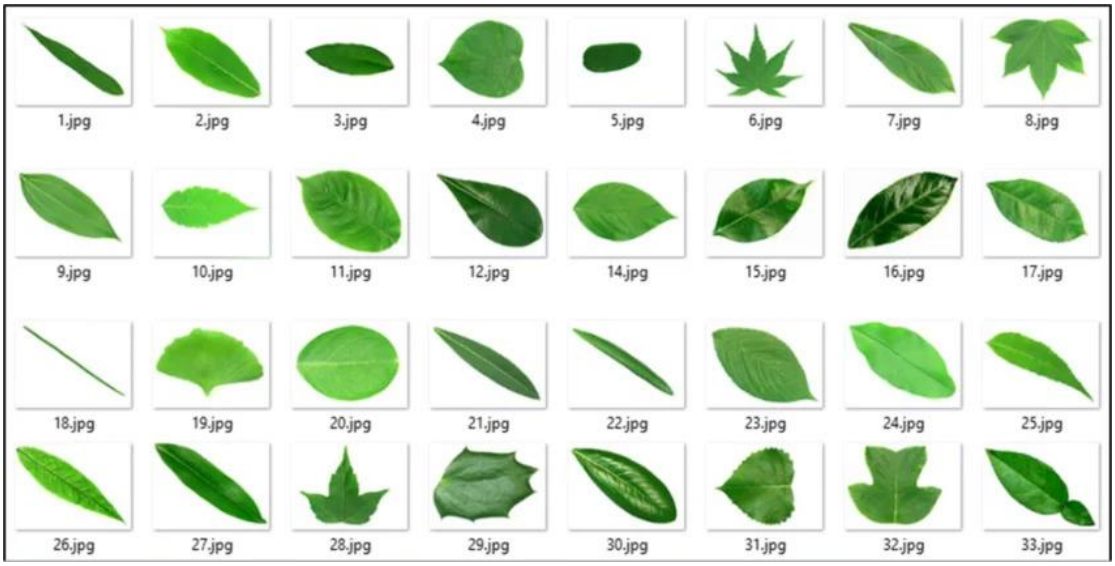


Fig 3. 5: Sample images from the FLAVIA dataset.

3.4.4. CIFAR-10 Dataset

The CIFAR-10 dataset contains 60,000 color images of size 32 × 32 pixels distributed over 10 classes, including natural object categories such as animals and vehicles. It was used to test the generalization capability of the proposed approach on complex image data [27].

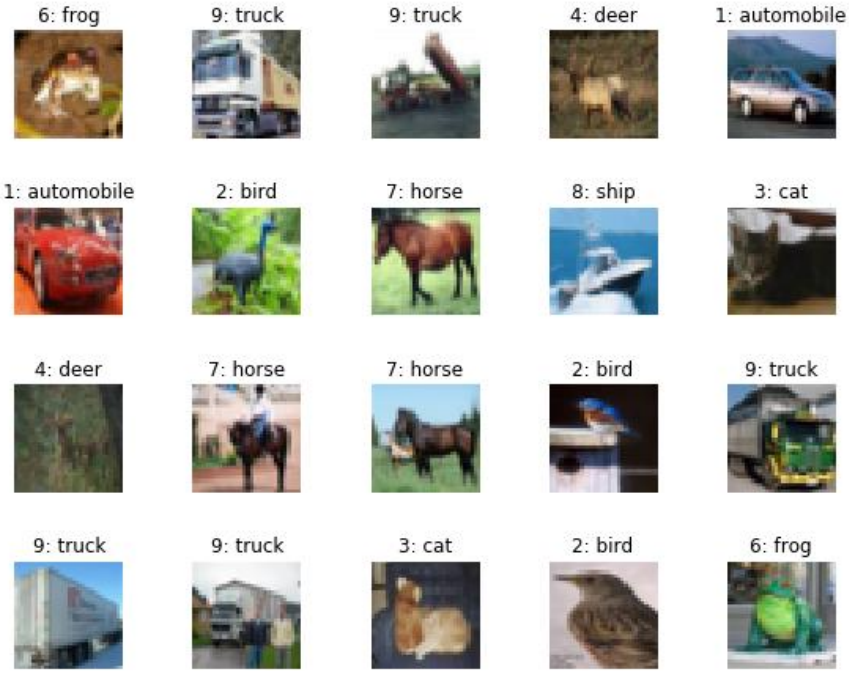


Fig 3. 6: Sample images from the CIFAR-10 dataset.

3.4.5. Four Shapes Dataset

The Four Shapes dataset contains 16,000 images distributed over four classes: circle, square, triangle, and star. Each image has a resolution of 200 × 200 pixels. This dataset was used to evaluate the effectiveness of moment descriptors under controlled geometric variations [28].

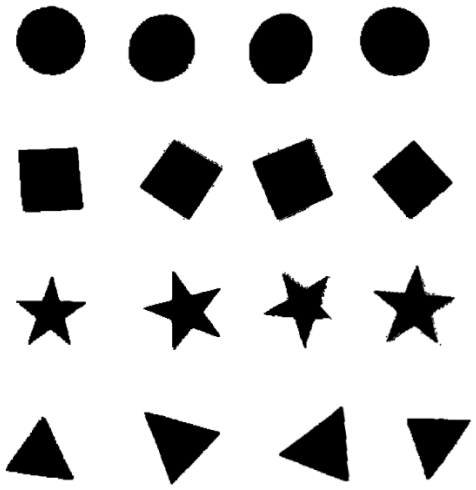


Fig 3. 7: Sample images from the Four Shapes dataset.

3.4.6. Vegetable Seed Image Dataset

The Vegetable Seed Image Dataset contains labeled images belonging to 14 seed classes and was included to assess the performance of the proposed system in agricultural image recognition applications [29].

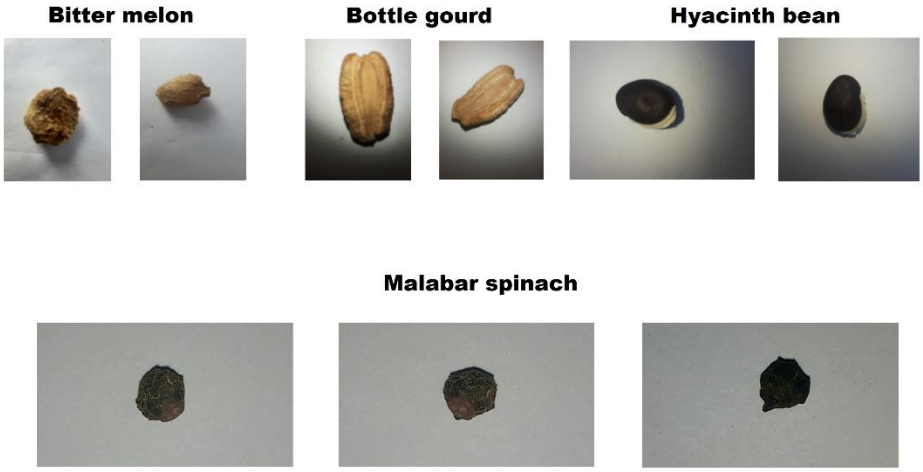


Fig 3. 8: Sample images from the Vegetable Seed Image dataset.

3.4.7. Malaria Cell Images Dataset

The Malaria Cell Images Dataset contains 27,558 microscopic images divided into two classes: Infected and Uninfected. It was used to evaluate descriptor performance in medical image classification tasks[30].

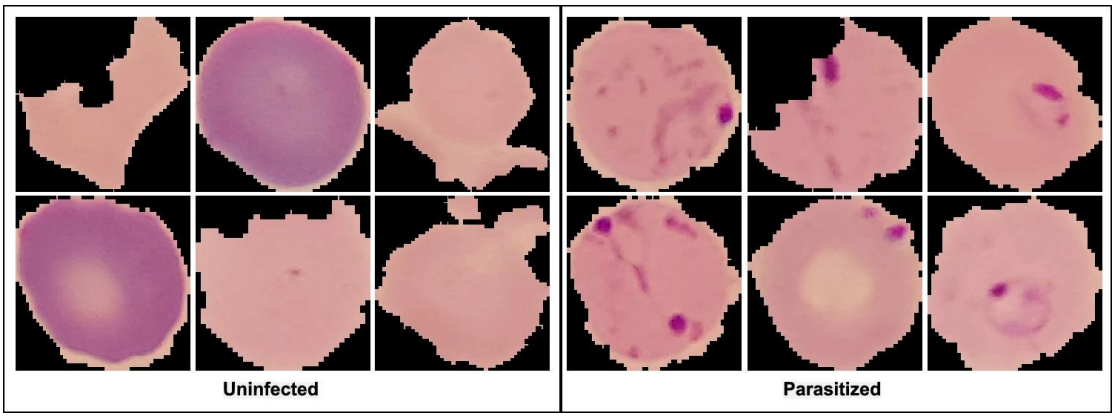


Fig 3. 9: Sample images from the Malaria Cell Images dataset.

3.4.8. GTSRB Dataset

The German Traffic Sign Recognition Benchmark contains more than 50,000 images and more than 40 traffic sign classes. It was used to evaluate the robustness of the proposed method under realistic visual variations[31].

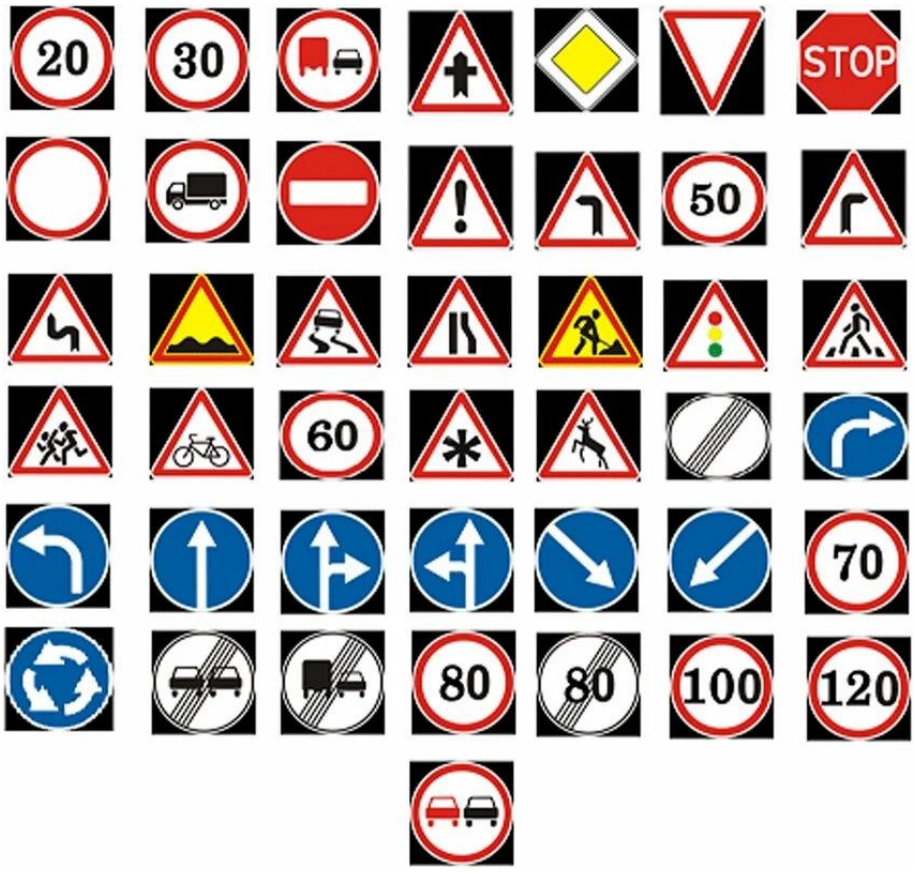


Fig 3. 10: Sample images from the GTSRB dataset.

3.4.9. Hand Gesture Recognition Dataset

The Hand Gesture Recognition dataset contains 10 gesture classes performed by 10 different subjects and was used to evaluate the performance of invariant moments in gesture recognition tasks[32].

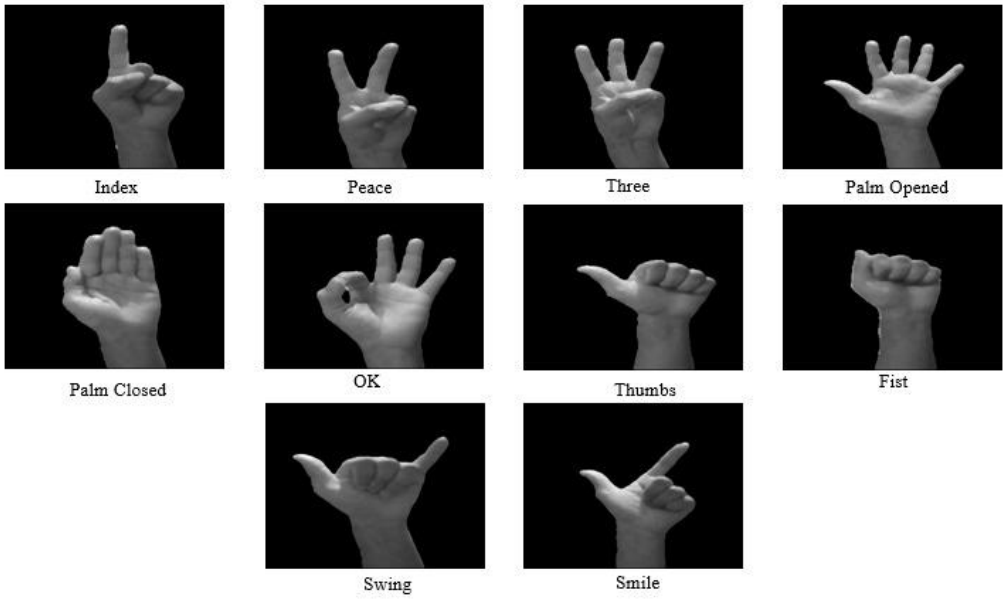


Fig 3. 11: Sample images from the Hand Gesture Recognition dataset.

3.4.10. Arabic Printed Digits Dataset

The Arabic Printed Digits Dataset was added to evaluate the effectiveness of invariant moment descriptors on printed digit recognition. The dataset is organized into ten folders, from digit_0 to digit_9, where each folder represents one digit class. Each class contains 624 binary images, resulting in a total of 6240 images.

The images were provided in a preprocessed form, after segmentation and binarization. Therefore, no additional segmentation or binarization step was applied in this experiment. The images were directly used for moment-based feature extraction after size standardization. This dataset is particularly suitable for evaluating the ability of invariant moments to describe isolated digit shapes in a controlled binary image recognition task[33].



Fig 3. 12: Sample images from the Arabic Printed Digits dataset.

3.5. Preprocessing

Before feature extraction, all input images were subjected to a preprocessing stage in order to improve their quality and prepare them for reliable moment computation. Since the proposed system is based on shape descriptors, preprocessing plays a crucial role in obtaining a clean, standardized, and consistent representation of the object. The main preprocessing operations applied in this work are described below.

3.5.1. Image Resizing

All input images were resized to a fixed resolution of 64 × 64 pixels. This operation ensures dimensional consistency across all datasets and reduces computational complexity during feature extraction. In addition, image resizing facilitates the comparison of descriptors by providing a uniform input representation for all experiments.

3.5.2. Grayscale Conversion

For color images, grayscale conversion was applied as a preliminary preprocessing step. This transformation reduces the image from multiple color channels to a single intensity channel while preserving the structural information required for shape analysis. Since moment descriptors primarily rely on object geometry rather than color information, grayscale conversion simplifies the data representation and improves computational efficiency.

3.5.3. Image Binarization

After grayscale conversion, image binarization was performed in order to separate the foreground object from the background. Gaussian smoothing was first applied to reduce small fluctuations and noise. Then, Otsu thresholding was used to automatically determine the optimal threshold value and generate a binary image suitable for shape extraction. This step is essential because moment descriptors require a clear and well-defined object representation.

3.5.4. Morphological Processing

Morphological operations were applied to refine the binary image and improve the quality of the extracted object. In particular, opening operations were used to remove small isolated artifacts, while closing operations were used to fill gaps and improve contour continuity. In addition, the largest connected component was retained in order to preserve the main object of interest and eliminate irrelevant regions.

3.5.5. Shape Normalization

The extracted object was normalized before feature computation. This process included centering the object, preserving its proportions, and positioning it consistently within a fixed image canvas. Shape normalization contributes directly to improving descriptor invariance and ensures that recognition performance is not affected by differences in object location or scale.

3.6. Feature Extraction Methods

After preprocessing, features were extracted using moment-based descriptors. These descriptors were selected because they provide compact numerical representations of image shape and offer invariance properties that are essential for recognition tasks.

3.6.1. Hu Moments

Hu moments are invariant descriptors derived from normalized central moments. They provide invariance to translation, rotation, and scaling and are widely used in shape recognition due to their simplicity and low computational cost.

3.6.2. Zernike Moments

Zernike moments are orthogonal moments defined on the unit circle. They provide compact and non-redundant feature representations and are known for their strong discriminative capability and robustness to rotational variations.

3.6.3. Pseudo-Zernike Moments

Pseudo-Zernike moments extend Zernike moments by providing a larger set of descriptors and improved robustness to noise. They are particularly effective in capturing fine shape variations.

3.6.4. Legendre Moments

Legendre moments are orthogonal polynomial-based descriptors characterized by efficient image representation and good approximation properties.

3.6.5. Tchebichef Moments

Tchebichef moments are discrete orthogonal moments particularly suitable for digital image analysis because they operate directly in the discrete domain and provide good numerical stability.

3.7. Experimental Setup

3.7.1. Dataset Splitting

Each dataset was divided into two subsets for training and testing purposes:

- ✓ Training set: 70%;
- ✓ Testing set: 30%.

The splitting process was performed using stratified sampling in order to preserve class distributions, while a fixed random state of 42 was used to ensure reproducibility.

3.7.2. Evaluation Metrics

The performance of the proposed image recognition system was evaluated using four standard classification metrics: accuracy, precision, recall, and F1-score. These metrics were computed from the confusion matrix, which summarizes the relationship between the predicted classes and the true classes [34].

In binary classification, the confusion matrix is generally based on four terms:

- True Positive (TP): samples correctly classified as positive;
- True Negative (TN): samples correctly classified as negative;
- False Positive (FP): negative samples incorrectly classified as positive;

- False Negative (FN): positive samples incorrectly classified as negative.

Although this work involves multi-class classification, these metrics can be computed for each class and then averaged to obtain global performance values.

A. Accuracy

Accuracy measures the proportion of correctly classified samples among the total number of samples. It provides a general indication of the overall performance of the classifier.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3.1)$$

A high accuracy value indicates that the classifier correctly predicts most of the samples. In this work, accuracy is the main metric used to compare the performance of invariant moment descriptors across different datasets, because it gives a direct measure of the global recognition rate.

B. Precision

Precision measures the proportion of correctly predicted positive samples among all samples predicted as positive. It evaluates the reliability of positive predictions made by the classifier.

$$Precision = \frac{TP}{TP+FP} \quad (3.2)$$

A high precision value means that the classifier produces few false positive predictions. In image recognition, precision is important when the objective is to reduce incorrect assignments of images to a given class.

C. Recall

Recall, also called sensitivity, measures the proportion of correctly identified positive samples among all actual positive samples. It evaluates the ability of the classifier to detect all samples belonging to a given class.

$$Recall = \frac{TP}{TP+FN} \quad (3.3)$$

A high recall value means that the classifier misses only a small number of samples from the target class. In image classification tasks, recall is important when it is necessary to correctly detect as many relevant images as possible.

D. F1-score

The F1-score is the harmonic mean of precision and recall. It provides a balanced measure between precision and recall, especially when there is an imbalance between classes.

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3.4)$$

A high F1-score indicates that the classifier achieves a good balance between avoiding false positives and reducing false negatives. Therefore, this metric is useful when accuracy alone is not sufficient to evaluate classification performance.

In this study, all four metrics were computed to provide a complete evaluation of the proposed recognition system. However, descriptor comparison is primarily based on classification accuracy, since it provides a direct and global measure of recognition performance across all datasets. Precision, recall, and F1-score were also used to better analyze the behavior of the classifier, especially in cases where some classes may be more difficult to recognize than others.

3.7.3. Classification in This Work

In this work, classification is performed using supervised learning techniques because all datasets used in the experiments are labeled. Each image belongs to a predefined class, such as a digit class in MNIST, a shape category in the Four Shapes dataset, a plant species in FLAVIA, or a gesture class in the Hand Gesture Recognition dataset.

Specifically, the k-Nearest Neighbors classifier is employed in the experimental phase. This choice is appropriate for the objective of this thesis because the main focus is not to design a complex classifier, but to evaluate the effectiveness of invariant moment descriptors as feature extraction methods.

The k-NN classifier assigns a class to a new sample by comparing its feature vector with the feature vectors of training samples and selecting the most frequent class among the nearest neighbors. This approach allows a direct evaluation of the discriminative power of Hu, Zernike, Pseudo-Zernike, Legendre, and Tchebichef moment descriptors.

Therefore, classification plays an essential role in the performance of the proposed recognition system. The final accuracy depends not only on the chosen classification algorithm, but also on the quality of preprocessing and feature extraction, especially when using shape-based descriptors such as invariant moments.

A. Implementation of the k-NN Classifier

The classification stage was implemented using the k-Nearest Neighbors algorithm. For each dataset and each moment descriptor, feature vectors were extracted from the preprocessed images and then divided into training and testing subsets.

During classification, each test sample was compared with the training samples in the feature space. The class label of the test sample was then determined according to the majority class among its nearest neighbors.

The use of k-NN is particularly suitable in this study because it provides a simple and transparent way to assess the separability of the extracted features. If a descriptor produces compact and discriminative feature vectors, samples from the same class are expected to be close to each other, leading to higher classification accuracy.

B. Distance Metric

In this work, Euclidean distance was used as the similarity measure for nearest-neighbor computation. It measures the distance between two feature vectors in the feature space.

A smaller Euclidean distance indicates greater similarity between two samples, while a larger distance indicates lower similarity. Therefore, the classification decision depends directly on the spatial distribution of feature vectors generated by each moment descriptor.

C. Feature Standardization

Before applying the k-NN classifier, feature standardization was performed using StandardScaler. This step is important because different moment descriptors may produce feature values with different numerical ranges.

Without standardization, features with larger numerical magnitudes could dominate the distance calculation and negatively affect the classification result. Standardization transforms the feature values so that they have comparable scales, which improves the reliability of the distance-based classification process.

D. Selection of k Values

Three values of k were investigated in the experiments:

k = 1;

k = 3;

k = 5.

For each descriptor and each dataset, the value of k that produced the highest classification accuracy was selected. This procedure allowed the classifier to adapt to the characteristics of the extracted features and the complexity of each dataset.

The comparison of different k values also helps to reduce the risk of selecting a classifier configuration that is too sensitive to individual samples. A smaller value of k may capture local class structures, while a larger value may provide more stable classification decisions.

3.8. Experimental Results

The proposed system was evaluated on ten datasets of different complexity levels in order to assess the effectiveness of the investigated invariant moment descriptors.

3.8.1. Results on MNIST

The MNIST dataset was used to evaluate the descriptors on simple handwritten digit recognition. The obtained accuracies are presented in Table 3.1.

Table 3. 1: Results on MNIST dataset

Method	**Accuracy (%)**
Legendre	94,47
Tchebichef	82,84
Zernike	84,21
Pseudo-Zernike	90,76
Hu	63,63

The results presented in Table 3.1 show clear differences in the performance of the studied moment descriptors on the MNIST dataset. Among all tested descriptors, Legendre moments achieved the highest classification accuracy of 94.47%, indicating superior effectiveness in representing handwritten digit shapes. This strong performance can be attributed to the orthogonality of Legendre moments and their ability to provide a compact and discriminative representation of global shape information.

Pseudo-Zernike moments ranked second with an accuracy of 90.76%, demonstrating good robustness and strong descriptive capability. Zernike moments and Tchebichef moments achieved comparable performance, with accuracies of 84.21% and 82.84%, respectively, showing that these descriptors remain effective for digit recognition, although they perform less efficiently than Legendre moments on this dataset.

Hu moments produced the lowest accuracy, 63.63%, significantly below the other descriptors. This lower performance suggests that the limited number of Hu invariant features and their reduced descriptive power may not be sufficient to capture subtle structural differences between handwritten digits.

Overall, the results indicate that orthogonal moments outperform Hu moments on the MNIST dataset, with Legendre moments providing the best recognition performance. These results confirm that more informative and less redundant descriptors can significantly improve classification accuracy for handwritten digit recognition.

3.8.2. Results on MPEG-7

The MPEG-7 dataset was used to assess the behavior of the descriptors on shape-based object recognition. The results are summarized in Table 3.2.

Table 3. 2: Results on MPEG-7 dataset

Method	**Accuracy (%)**
Legendre	81,42
Tchebichef	71,41
Zernike	76,72
Pseudo-Zernike	76,00
Hu	57,90

The results presented in Table 3.2 show differences in the performance of the studied moment descriptors on the MPEG-7 dataset. Among all tested descriptors, Legendre moments achieved the highest classification accuracy of 81.42%, indicating the best performance for shape-based object recognition in this dataset. This result suggests that Legendre moments

provide an effective representation of global shape characteristics and good discriminative capability.

Zernike moments and Pseudo-Zernike moments achieved similar performance, with accuracies of 76.72% and 76.00%, respectively. These results confirm that both descriptors are effective for shape recognition and provide reliable feature representations.

Tchebichef moments obtained an accuracy of 71.41%, showing acceptable performance, although lower than Legendre, Zernike, and Pseudo-Zernike moments.

Hu moments produced the lowest accuracy, 57.90%, indicating weaker performance compared to the other descriptors. This result suggests that the limited descriptive capability of Hu moments may reduce their effectiveness in representing complex object shapes.

Overall, the results indicate that Legendre moments provide the best recognition performance on the MPEG-7 dataset, while orthogonal moments generally outperform Hu moments.

3.8.3. Results on Four Shapes

The Four Shapes dataset contains synthetic geometric images of circles, squares, triangles, and stars. The results are shown in Table 3.3.

Table 3. 3: Results on Four Shapes dataset

Method	**Accuracy (%)**
Legendre	100,00
Tchebichef	100,00
Zernike	100,00
Pseudo-Zernike	100,00
Hu	95,70

The results presented in Table 3.3 show excellent classification performance for all studied moment descriptors on the Four Shapes dataset. Legendre, Tchebichef, Zernike, and Pseudo-Zernike moments all achieved perfect classification accuracy of 100%, indicating that these descriptors were able to fully discriminate between the four geometric shape classes.

This result suggests that these descriptors provide highly effective representations for simple and well-defined geometric objects, where inter-class differences are sufficiently distinct to allow complete separation in the feature space.

Hu moments also achieved a high accuracy of 95.70%, although slightly lower than the other descriptors. This result indicates that Hu moments remain effective for this dataset, but their descriptive capability is comparatively less accurate than the orthogonal moments.

Overall, the results indicate that all studied descriptors perform very well on the Four Shapes dataset, with orthogonal moments achieving perfect recognition performance.

3.8.4. Results on FLAVIA

The FLAVIA dataset was used to evaluate the proposed Method on leaf recognition. The results are reported in Table 3.4.

Table 3. 4: Results on FLAVIA dataset

Method	**Accuracy (%)**
Legendre	76,96
Tchebichef	63,87
Zernike	71,20
Pseudo-Zernike	70,16
Hu	48,69

The results presented in Table 3.4 show differences in the performance of the studied moment descriptors on the FLAVIA dataset. Among all tested descriptors, Legendre moments achieved the highest classification accuracy of 76.96%, indicating the best performance for leaf recognition in this dataset. This result suggests that Legendre moments provide effective representation of the shape characteristics of leaf images.

Zernike moments achieved an accuracy of 71.20%, while Pseudo-Zernike moments obtained a close performance of 70.16%, showing that both descriptors provide reliable feature representations for this recognition task.

Tchebichef moments reached an accuracy of 63.87%, indicating acceptable performance, although lower than Legendre, Zernike, and Pseudo-Zernike moments.

Hu moments produced the lowest accuracy, 48.69%, indicating weaker performance compared to the other descriptors. This result suggests that Hu moments may have limited effectiveness in representing the structural complexity of leaf shapes.

Overall, the results indicate that Legendre moments provide the best recognition performance on the FLAVIA dataset, while orthogonal moments generally outperform Hu moments.

3.8.5. Results on CIFAR-10

The CIFAR-10 dataset was used to examine the limitations of shape-based invariant moments on complex natural images. The results are presented in Table 3.5.

Table 3. 5: Results on CIFAR-10 dataset

Method	**Accuracy (%)**
Legendre	27,03
Tchebichef	20,37
Zernike	19,50
Pseudo-Zernike	19,72
Hu	14,65

The results presented in Table 3.5 show relatively low classification performance for all studied moment descriptors on the CIFAR-10 dataset. Among all tested descriptors, Legendre moments achieved the highest classification accuracy of 27.03%, indicating the best performance within this dataset.

Tchebichef moments obtained an accuracy of 20.37%, while Pseudo-Zernike and Zernike moments achieved similar results, with accuracies of 19.72% and 19.50%, respectively. These results indicate comparable performance among these descriptors.

Hu moments produced the lowest accuracy, 14.65%, indicating weaker performance compared to the other descriptors.

Overall, the results indicate that Legendre moments provide the best recognition performance on the CIFAR-10 dataset, while all descriptors show limited effectiveness for this recognition task.

3.8.6. Results on Vegetable Seeds

The Vegetable Seed dataset was used as a practical real-world biological application. The obtained results are shown in Table 3.6.

Table 3. 6: Results on Vegetable Seeds dataset

Method	**Accuracy (%)**
Legendre	68,87
Tchebichef	63,85
Zernike	66,87
Pseudo-Zernike	67,69
Hu	55,64

The results presented in Table 3.6 show differences in the performance of the studied moment descriptors on the Vegetable Seed dataset. Among all tested descriptors, Legendre moments achieved the highest classification accuracy of 68.87%, indicating the best recognition performance in this dataset.

Pseudo-Zernike moments obtained an accuracy of 67.69%, followed closely by Zernike moments with 66.87%, showing that these descriptors provide similar and reliable performance for this recognition task.

Tchebichef moments achieved an accuracy of 63.85%, indicating acceptable performance, although lower than Legendre, Pseudo-Zernike, and Zernike moments.

Hu moments produced the lowest accuracy, 55.64%, indicating weaker performance compared to the other descriptors.

Overall, the results indicate that Legendre moments provide the best recognition performance on the Vegetable Seed dataset, while orthogonal moments generally outperform Hu moments.

3.8.7. Results on Malaria Cell Images

The Malaria Cell Images dataset was used to evaluate the performance of the proposed descriptors on a medical image classification task. The obtained results are presented in Table 3.7.

Table 3. 7: Results on Malaria Cell Images

Method	**Accuracy (%)**
Legendre	58,61
Tchebichef	57,24
Zernike	60,80
Pseudo-Zernike	60,41
Hu	58,76

The results presented in Table 3.7 show relatively close performance among the studied moment descriptors on the Malaria Cell Images dataset. Among all tested descriptors, Zernike moments achieved the highest classification accuracy of 60.80%, indicating the best recognition performance in this dataset.

Pseudo-Zernike moments obtained an accuracy of 60.41%, showing very close performance to Zernike moments. Hu moments and Legendre moments also achieved similar results, with accuracies of 58.76% and 58.61%, respectively.

Tchebichef moments produced the lowest accuracy, 57.24%, although the performance difference among all descriptors remains relatively small.

Overall, the results indicate that the studied descriptors show comparable performance on the Malaria Cell Images dataset, with Zernike moments providing the best recognition performance.

3.8.8. Results on German Traffic Sign Recognition Benchmark

The German Traffic Sign Recognition Benchmark dataset was used to evaluate the performance of the proposed descriptors on a traffic sign classification task. The obtained results are presented in Table 3.8.

Table 3. 8: Results on German Traffic Sign Recognition Benchmark

Method	**Accuracy (%)**
Legendre	45,13
Tchebichef	22,20
Zernike	35,58
Pseudo-Zernike	32,39
Hu	17,70

The results presented in Table 3.8 show differences in the performance of the studied moment descriptors on the German Traffic Sign Recognition Benchmark dataset. Among all tested descriptors, Legendre moments achieved the highest classification accuracy of 45.13%, indicating the best recognition performance in this dataset.

Zernike moments obtained an accuracy of 35.58%, while Pseudo-Zernike moments achieved 32.39%, showing acceptable performance, although lower than Legendre moments.

Tchebichef moments reached an accuracy of 22.20%, indicating lower effectiveness compared to the other orthogonal descriptors. Hu moments produced the lowest accuracy, 17.70%, showing the weakest performance among all tested descriptors.

Overall, the results indicate that Legendre moments provide the best recognition performance on the German Traffic Sign Recognition Benchmark dataset, while orthogonal moments generally outperform Hu moments.

3.8.9. Results on Hand Gesture Recognition Dataset

The Hand Gesture Recognition dataset was used to evaluate the performance of the proposed descriptors on a gesture classification task. The obtained results are presented in Table 3.9.

Table 3. 9: Results on Hand Gesture Recognition Dataset

Method	**Accuracy (%)**
Legendre	99,90
Tchebichef	99,03
Zernike	99,42
Pseudo-Zernike	99,72
Hu	96,92

The results presented in Table 3.9 show excellent classification performance for all studied moment descriptors on the Hand Gesture Recognition dataset. Among all tested descriptors, Legendre moments achieved the highest classification accuracy of 99.90%, indicating the best recognition performance in this dataset.

Pseudo-Zernike moments achieved an accuracy of 99.72%, while Zernike and Tchebichef moments obtained 99.42% and 99.03%, respectively, showing that all orthogonal descriptors provide highly effective feature representations for this recognition task.

Hu moments also achieved a high accuracy of 96.92%, although slightly lower than the other descriptors.

Overall, the results indicate that all studied descriptors perform very well on the Hand Gesture Recognition dataset, with Legendre moments providing the best recognition performance.

3.8.10. Results on Arabic Printed Digits Dataset

The Arabic Printed Digits Dataset was used to evaluate the performance of the studied invariant moment descriptors on printed digit classification. The obtained results are presented in Table 3.10.

Table 3. 10: Results on Arabic Printed Digits Dataset

Method	**Accuracy (%)**
Legendre	95,19
Tchebichef	82,91
Zernike	86,38
Pseudo-Zernike	84,19
Hu	70,94

The results show the classification performance of the studied moment descriptors on the Arabic printed digit dataset. Since the images were already segmented and binarized, the experiment focused mainly on the ability of each descriptor to represent the geometric structure of printed digits. The comparison between Hu, Zernike, Pseudo-Zernike, Legendre, and Tchebichef moments allows evaluating which descriptor provides the most discriminative representation for this type of binary digit recognition task.

3.9. Global Comparison

Table 3. 11: Overall Comparative Performance of Invariant Moment Descriptors

Descriptor	**Recognition Capability**	**Robustness**	**Overall Performance**
Legendre	Excellent	High	Best overall
Zernike	Very Good	High	Strong
Pseudo-Zernike	Very Good	High	Strong
Tchebichef	Good	Moderate	Acceptable
Hu	Moderate	Lower	Limited

The results obtained across all datasets confirm that invariant moments are effective descriptors for image recognition and can successfully represent shape information for classification tasks.

The experimental results show that all studied descriptors were able to perform image recognition with varying levels of effectiveness. In particular, orthogonal moments such as Legendre, Zernike, and Pseudo-Zernike demonstrated strong recognition capability and robust performance across multiple datasets.

Among the studied descriptors, Legendre moments provided the most consistent and effective overall performance. Zernike and Pseudo-Zernike moments also showed strong recognition capability, while Tchebichef moments produced acceptable performance. Hu moments generally yielded lower accuracies but still demonstrated the ability to perform image recognition.

The results also show that invariant moments are particularly effective for shape-oriented recognition tasks, where very high accuracies were achieved, while their performance becomes more limited for highly complex natural image datasets.

Overall, the global comparison confirms not only differences between descriptors, but also the effectiveness of invariant moments as practical tools for image recognition.

3.10. Comparison with Previous Studies

This section compares the obtained results with previous studies that used the same datasets. The objective of this comparison is to position the results of the present work with respect to existing research and to highlight the influence of descriptors, classifiers, preprocessing strategies, and experimental settings on recognition accuracy. Only studies conducted on datasets with the same name are considered in order to ensure a fair and meaningful comparison.

3.10.1. MNIST Dataset

After presenting the results obtained on the MNIST dataset, our findings are compared with those reported by Bourzik et al. (2024). Their study investigated invariant Tchebichef moments and invariant Krawtchouk moments for handwritten digit recognition. Table 3.12 summarizes the comparison between their reported results and the results obtained in the present work [35].

Table 3. 12: Comparison with Previous Studies on MNIST Dataset

Study	Descriptor / Method	Classifier	Reported Accuracy (%)	Comparison with Our Result
Bourzik et al. (2024)	Invariant Tchebichef Moments (TMI)	1-NN	90.60	Higher than our Tchebichef result
Bourzik et al. (2024)	Invariant Tchebichef Moments (TMI)	Neural Network	57.40	Lower than our Tchebichef result
Bourzik et al. (2024)	Invariant Krawtchouk Moments (KMI)	1-NN	99.30	Higher, but different descriptor
Bourzik et al. (2024)	CNN baseline	CNN	82.80	Very close to our Tchebichef result
Our work	Legendre Moments	k-NN	94.47	Best result in our MNIST experiments; higher than Bourzik et al.'s Tchebichef result and CNN baseline
Our work	Tchebichef Moments	k-NN	82.84	Current experimental result

A. Analysis

The study by Bourzik et al. (2024) compared Invariant Tchebichef Moments (TMI) and Invariant Krawtchouk Moments (KMI) for handwritten digit recognition on a reduced version of the MNIST dataset. Their results showed an accuracy of 90.60% using 1-NN with Tchebichef Moments, 99.30% using 1-NN with Krawtchouk Moments, and 82.80% using a CNN baseline.

In the present work, the Tchebichef moments achieved an accuracy of 82.84% on the MNIST dataset, which is very close to the CNN baseline reported by Bourzik et al. (2024). However, the best result in our MNIST experiments was obtained using Legendre moments, with an accuracy of 94.47%. This result is higher than the accuracy reported by Bourzik et al. for Tchebichef Moments and also higher than their CNN baseline.

Although Bourzik et al. reported a higher accuracy of 99.30% using Krawtchouk Moments, this descriptor was not used in the present work. Therefore, the comparison mainly shows that Legendre moments provided the strongest performance among the descriptors investigated in this thesis, while Tchebichef moments remained competitive when compared with the CNN baseline.

The difference in results may be explained by several factors, including the descriptor used, the classifier configuration, the dataset size, and the preprocessing strategy. Overall, this comparison confirms that invariant moments can provide competitive performance for

handwritten digit recognition, especially when an appropriate descriptor such as Legendre moments is used.

B. Discussion

The comparison highlights that Invariant Moments, particularly Tchebichef Moments, are effective descriptors for handwritten digit recognition, especially when paired with k-NN. While Krawtchouk Moments achieved higher accuracy, this suggests that Krawtchouk Moments could be explored in future studies to enhance performance further. However, CNN results in their study were lower than both Krawtchouk and Tchebichef Moments, indicating that Invariant Moments can be a viable and efficient alternative to CNN for certain types of recognition tasks.

The results also underscore the importance of feature extraction methods and classifiers in determining the performance of image recognition systems. While CNN is a powerful tool, Invariant Moments offer a computationally efficient approach with competitive performance, especially for simple shape recognition tasks.

3.10.2. Four Shapes Dataset

In addition to the results obtained in the present work, a previous experiment on the same Four Shapes dataset reported very high recognition performance using Hu Moments. Since this dataset contains simple and well-defined geometric classes, namely circle, square, triangle, and star, it represents a suitable benchmark for evaluating the effectiveness of shape-based moment descriptors. The following table compares the reported result of the previous study with the results obtained in our work[36].

Table 3. 13: Comparison with Previous Studies on Four Shapes Dataset

Study	Descriptor / Method	Classifier	Reported Accuracy (%)	Comparison with Our Result
Hu Moments (2024)	Hu Moments	SVM	99.98	Higher than our result
Our work	Tchebichef Moments	k-NN	100.00	Current experimental result

A. Analysis

The study "Four Shapes using Hu Moments" achieved 99.98% accuracy using Hu Moments for shape recognition on the Four Shapes dataset, with a very low standard deviation of 0.02. This is very close to the 100% accuracy we achieved using Tchebichef Moments in our study.

The difference in the reported accuracy could be attributed to the specific moments used (Hu Moments vs. Tchebichef Moments). While both descriptors are highly effective for shape recognition, Tchebichef Moments performed slightly better in our study, achieving 100% accuracy compared to 99.98% with Hu Moments. This indicates that Tchebichef Moments may have a slight advantage in representing the shapes within this dataset, leading to perfect classification in our experiment.

B. Discussion

This comparison highlights the strong performance of both Hu Moments and Tchebichef Moments for shape recognition tasks. Hu Moments are known for their simplicity and effectiveness, achieving nearly perfect accuracy in the study, while Tchebichef Moments provided perfect accuracy in our experiments. The slight difference in performance could suggest that Tchebichef Moments might offer a better representation for this dataset.

Both approaches demonstrate the power of Invariant Moments for shape-based recognition, and their performance is highly dependent on the dataset and the feature extraction method used. Future studies could explore the combination of both methods to further improve accuracy in various recognition tasks.

3.10.3. MPEG-7 Dataset

In this section, we compare the results obtained in our study with the study by Kalami and Sedghi (2019), who also used the MPEG-7 dataset for shape classification. Their study reported results using Rotation Invariant Moments and Generated Stationary Transformed Features, while our study focused on Invariant Moments for classification tasks. Below is a comparison of the reported results and our experimental findings.

Table 3. 14: Comparison with Previous Studies on MPEG-7 Dataset

Study	Descriptor / Method	Classifier / Evaluation	Reported Result	Comparison with Our Result
Kalami and Sedghi (2019)	Rotation Invariant Moments and Generated Stationary Transformed Features	Retrieval Evaluation (Average Recall, Average Precision)	Average Recall ~0.57; Average Precision ~0.93	Not directly comparable, as retrieval metrics were used
Kalami and Sedghi (2019)	Zernike Moments	Retrieval Evaluation (Average Recall,	Average Recall ~0.50; Average	Different descriptor and evaluation method

		Average Precision)	Precision ~0.83	
Our work	Legendre Moments	k-NN classification	81,42% accuracy	Best result in our MPEG-7 experiments; direct comparison with Kalami and Sedghi is limited because different evaluation metrics were used.
Our work	Zernike Moments	k-NN classification	76,72% accuracy	Second-best result in our MPEG-7 experiments; not directly comparable with Kalami and Sedghi due to different evaluation metrics.
Our work	Pseudo-Zernike Moments	k-NN classification	76,00% accuracy	Very close to Zernike Moments in our work, with slightly lower accuracy.
Our work	Tchebichef Moments	k-NN classification	71.41% accuracy	Acceptable performance, but lower than Legendre, Zernike, and Pseudo-Zernike Moments in our work.
Our work	Hu Moments	k-NN classification	57.90% accuracy	Acceptable performance, but lower than Legendre, Zernike, and Pseudo-Zernike Moments in our work.

A. Analysis

The study by Kalami and Sedghi (2019) [37] primarily focused on Rotation Invariant Moments and Generated Stationary Transformed Features for shape retrieval on the MPEG-7 dataset. Their results showed high retrieval performance, particularly in terms of Average Precision. They also compared their proposed method with other descriptors such as Zernike Moments and Shape Signature using retrieval-based evaluation metrics, including Average Recall and Average Precision.

In comparison, the present study used a supervised classification approach based on the k-Nearest Neighbors classifier with several invariant moment descriptors, namely Legendre Moments, Zernike Moments, Pseudo-Zernike Moments, Tchebichef Moments, and Hu Moments. The best result in our MPEG-7 experiments was achieved using Legendre Moments, with an accuracy of 81.42%. This indicates that Legendre Moments provided the most discriminative representation for shape-based classification in this dataset.

Zernike Moments and Pseudo-Zernike Moments achieved close results, with accuracies of 76.72% and 76.00%, respectively. This confirms that both descriptors are effective for representing object shapes in the MPEG-7 dataset. Tchebichef Moments obtained an accuracy of 71.41%, showing acceptable performance, although lower than Legendre, Zernike, and Pseudo-Zernike Moments. Hu Moments produced the lowest accuracy, with 57.90%, which suggests that their limited descriptive capacity may be less suitable for representing complex shape variations in MPEG-7.

However, a direct numerical comparison with Kalami and Sedghi (2019) is limited because their study used retrieval metrics, whereas the present work used classification accuracy. Therefore, the comparison should be interpreted mainly as a methodological positioning rather than a direct performance ranking.

B. Discussion

The comparison with Kalami and Sedghi (2019) shows that invariant moments remain relevant descriptors for MPEG-7 shape analysis. Nevertheless, the comparison is not fully direct because the two studies rely on different evaluation protocols. Kalami and Sedghi evaluated shape retrieval using Average Recall and Average Precision, whereas the present study evaluated supervised classification accuracy using k-NN.

Despite this methodological difference, the results obtained in this work confirm that invariant moment descriptors can effectively represent object shapes for classification tasks. Among the studied descriptors, Legendre Moments achieved the best performance on MPEG-7, followed by Zernike and Pseudo-Zernike Moments. This suggests that orthogonal moments provide more discriminative and stable shape representations than conventional Hu Moments.

The lower performance of Hu Moments indicates that simple invariant descriptors may not be sufficient to capture the complex shape variations present in the MPEG-7 dataset. Therefore, future work could explore the combination of several moment descriptors, the integration of additional shape features, or the use of more advanced classifiers in order to improve recognition performance.

3.11. Discussion

The obtained results confirm that invariant moments constitute effective descriptors for image recognition and can successfully represent image shape information for classification tasks.

The experimental results showed that the studied moment descriptors were capable of recognizing image classes across different datasets and application domains, including handwritten digits, biological images, medical images, traffic signs, gesture recognition, and Arabic printed digits.

The strong performance observed for orthogonal moments, particularly Legendre, Zernike, and Pseudo-Zernike moments, confirms that invariant moments provide robust and discriminative shape representations. Their invariance properties contribute significantly to recognition performance under geometric transformations.

In particular, **Arabic Printed Digits Dataset** demonstrated the effectiveness of invariant moments in recognizing printed digit shapes. Legendre moments achieved the highest performance on this dataset, with an accuracy of 95.19%, confirming that they are highly effective for digit recognition tasks.

The weaker performance of Hu moments in several experiments does not invalidate the usefulness of invariant moments, but rather highlights differences in descriptive power among descriptor families. This was particularly evident in the **Arabic Printed Digits Dataset**, where Hu moments showed the lowest accuracy at 70.94%.

The results also show that the effectiveness of invariant moments depends on the nature of the recognition task. For shape-oriented datasets, such as the Arabic **Printed Digits Dataset** and the **Four Shapes Dataset**, invariant moments provide excellent recognition performance, while for more complex natural images, their effectiveness may become more limited.

These observations are consistent with the theoretical analysis presented in Chapter 2 and confirm that invariant moments are not only mathematical descriptors, but also practical tools for image recognition applications, particularly in well-defined shape recognition tasks.

3.12. Conclusion

In this chapter, the proposed recognition system was implemented and evaluated using multiple datasets and five different moment descriptors. The experimental results confirmed that invariant moments can be effectively used for image recognition and classification tasks.

The obtained results showed that the studied descriptors were capable of recognizing image classes with varying levels of performance depending on the dataset characteristics. Among them, Legendre moments provided the best overall performance in most datasets, while Zernike

and Pseudo-Zernike moments also showed strong effectiveness. Hu moments generally produced the lowest performance.

The global comparison and discussion also confirmed that orthogonal moments provide more effective shape representations and better recognition performance than conventional invariant moments.

Overall, these results validate the effectiveness of the proposed approach and demonstrate that invariant moments constitute useful and practical tools for image recognition applications.

Conclusion

In this work, image recognition using invariant moments was studied through the implementation and evaluation of several moment-based descriptors on multiple datasets and recognition tasks.

The obtained results demonstrate that invariant moments are effective tools for image recognition and provide meaningful shape representations that can be successfully used for classification. The experimental results confirm that moment descriptors are capable of recognizing and discriminating image classes with good performance, particularly in tasks where shape information plays a significant role.

This study also showed that orthogonal moments, especially Legendre, Zernike, and Pseudo-Zernike moments, provide strong representation capabilities and can improve recognition performance due to their invariance properties and discriminative power.

Beyond classification performance, this work demonstrates that invariant moments have practical value in several real-world applications. In biological sciences, they can support plant recognition and seed classification. In medical applications, they can contribute to microscopic image analysis and computer-aided diagnosis. In intelligent systems, they can be applied to gesture recognition, traffic sign recognition, and other automated recognition tasks.

These applications may also be beneficial for students and researchers in biology, medicine, agriculture, and engineering, as they provide practical tools for image analysis and intelligent decision support.

Therefore, this work confirms not only that invariant moments are useful for image recognition, but also that they have practical relevance in education, research, and daily-life applications.

Despite the promising results obtained, some limitations remain, particularly for highly complex image recognition tasks. As future work, this study can be extended by combining invariant moments with deep learning methods, hybrid feature extraction approaches, or more advanced classifiers to further improve recognition performance.

References

[1] R. C. Gonzalez and R. E. Woods, “Digital Image Processing,” 4th ed., New York: Pearson, 2018.

[2] R. Szeliski, "Computer vision: Algorithms and applications," 2nd ed. Springer, 2022.

[3] D. Zhang, "Image processing and analysis: Variational, PDE, wavelet, and stochastic methods," Springer, 2012.

[4] M. K. Hu, “Visual pattern recognition by moment invariants,” *IRE Transactions on Information Theory*, vol. 8, no. 2, pp. 179–187, 1962.

[5] R. C. Gonzalez and R. E. Woods, “Digital Image Processing,” 4th ed., New York: Pearson, 2018.

[6] R. Szeliski, Computer Vision: Algorithms and Applications, 2nd ed. Springer, 2022.

[7] R. C. Gonzalez and R. E. Woods, “Digital Image Processing,” 4th ed., New York: Pearson, 2018.

[8] R. C. Gonzalez and R. E. Woods, “Digital Image Processing,” 4th ed., New York: Pearson, 2018.

[9] D. Zhang, Image Processing and Analysis. Berlin, Germany: Springer, 2012.

[10] Stanford University, "Image Recognition Tutorial,"
Accessed on: Apr. 5, 2026. [Online].
Available: https://ai.stanford.edu/~syyeung/cvweb/tutorial4.html

[11] C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer, 2006.

[12] T. Cover and P. Hart, “Nearest neighbor pattern classification,” IEEE Transactions on Information Theory, vol. 13, no. 1, pp. 21–27, 1967.

[13] C. Cortes and V. Vapnik, “Support-vector networks,” Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.

[14] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, Classification and Regression Trees. Belmont, CA, USA: Wadsworth, 1984.

[15] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.

[16] M. K. Hu, “Visual pattern recognition by moment invariants,” 1962.

[17] R. C. Gonzalez and R. E. Woods, “Digital Image Processing,” 4th ed., New York: Pearson, 2018.

[18] S. X. Liao, “Image Analysis by Moments,” PhD Thesis, University of Manitoba, Canada, 2003.

[19] M. K. Hu, “Visual pattern recognition by moment invariants,” IRE Transactions on Information Theory, vol. 8, no. 2, pp. 179–187, 1962.

[20] J. Flusser, B. Zitová, and T. Suk, “Moments and Moment Invariants in Pattern Recognition,” Chichester, UK: Wiley, 2009.

[21] A. Khotanzad and Y. H. Hong, “Invariant image recognition by Zernike moments,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 5, pp. 489–497, 1990.

[22] R. Mukundan and K. R. Ramakrishnan, “Moment Functions in Image Analysis: Theory and Applications,” Singapore: World Scientific, 1998.

[23] R. Mukundan, “Some computational aspects of discrete orthonormal moments,” IEEE Transactions on Image Processing, vol. 13, no. 8, pp. 1055–1059, 2004.

[24] "MNIST Dataset", Kaggle. Accessed: Mar. 19, 2026. [Online]. Available: https://www.kaggle.com/datasets/hojjatk/mnist-dataset

[25] "MPEG-7 Shape Dataset", Temple University. Accessed: Mar. 25, 2026. [Online]. Available: https://dabi.temple.edu/external/shape/MPEG7/dataset.html

[26] "Flavia_dataset", Kaggle. Accessed: Mar. 28, 2026. [Online]. Available: https://www.kaggle.com/datasets/gauravneupane/flavia-dataset

[27] "CIFAR-10 and CIFAR-100 datasets", University of Toronto. Accessed: Apr. 02, 2026. [Online]. Available: https://www.cs.toronto.edu/~kriz/cifar.html

[28] "Four Shapes Dataset", Kaggle. Accessed: Apr. 04, 2026. [Online]. Available: https://www.kaggle.com/datasets/smeschke/four-shapes

[29] "Vegetable Seed Dataset", Kaggle. Accessed: Apr. 09, 2026. [Online]. Available: https://www.kaggle.com/datasets/imr4n4lif/vegetable-seed-dataset

[30] "Malaria Cell Images Dataset ", Kaggle. Accessed: Apr. 10, 2026. [Online]. Available: https://www.kaggle.com/datasets/iarunava/cell-images-for-detecting-malaria

[31] "GTSRB - German Traffic Sign Recognition Benchmark", Kaggle. Accessed: Apr. 11, 2026. [Online]. Available: https://www.kaggle.com/datasets/meowmeowmeowmeowmeow/gtsrb-german-traffic-sign

[32] " Multi-Modal Dataset for Hand Gesture Recognition", Kaggle. Accessed: Apr. 22, 2026. [Online]. Available: https://www.kaggle.com/datasets/gti-upm/multimodhandgestrec

[33] " Printed arabic digits ", Kaggle. Accessed: Apr. 29, 2026. [Online]. Available: https://www.kaggle.com/datasets/derdour/printed-arabic-digits/

[34] M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” Information Processing & Management, vol. 45, no. 4, pp. 427–437, 2009.

[35] A. Bourzik, B. Bouikhalene, J. El-Mekkaoui, and A. Hjouji, “Invariant Tchebichef and Krawtchouk Moments for Shape Recognition: A Performance Evaluation Study,” in Proceedings of the 7th International Conference on Networking, Intelligent Systems and Security (NISS 2024), 2024. doi: 10.1145/3659677.3659709.

[36] "Four Shapes using Hu Moments", Kaggle. Accessed: Apr. 30, 2026. [Online]. Available: https://www.kaggle.com/code/keizerzilla/four-shapes-using-hu-moments-99-98-acc-0-02-std

[37] "Database Classification of MPEG7 Shape Using Rotation Moments and Generated Stationary Transformed Features ", Current Journal of Applied Science and Technology. Accessed: Apr. 29, 2026. [Online]. Available: https://journalcjast.com/index.php/CJAST/article/view/2078