Title: Evaluate the performance of PCA, SIFT and HOG features in object recognition tasks for industrial automation.

Author: Vineet Muley- 229310401

Abstract

This case study examines the performance of three prominent characteristic extraction techniques (main component analysis), scale-invariant feature transformation (SIFT) and histogram-oriented gradient (HOG) for industrial automation object detection tasks. With the COIL-20 dataset, all technology is applied to extract features from grey landscape images from 20 different objects and classify them into KNN classifiers (K-Nearest Neighbours). PCA and HOG achieved both high classification accuracy of 98.96% and 98.26%. This demonstrates its effectiveness in extracting robust features for object classification. In contrast, sieving was performed in conjunction with a visual word bag (BOVW), resulting in a 87.5% lower accuracy, highlighting the potential limitations of displaying properties on tactile-based features in even vector rooms. Standard metrics including accuracy, accuracy, recall and F1 score. This study provides comparative insight into the applicability and performance of various characteristic extraction methods, and provides guidelines for selecting characteristics in real industrial vision systems.

Introduction

1. **Background**

   Object detection plays an important role in industrial automation, allowing machines to identify, classify and interact with physical objects in manufacturing and inspection environments. Computer vision technology is becoming increasingly important in modern industrial systems due to its ability to reduce human intervention, improve accuracy and accelerate quality control processes. The basis for successful object recognition lies in effective property extraction, including the transformation of raw pixel data into meaningful representations that can interpret machine learning algorithms. In a variety of distinctive extraction methods, Main Component Analysis (PCA), scaling invariance is widespread due to its reliability and validity. Each of these techniques records different aspects of the PCA image, reduces dimensions, and variance shows record of talent and local descriptor records, as well as pigs and object structures. This study focuses on comparing these three methods to assess performance in an industrial context with the help of standardized data records.

2. Problem Statement

Despite the rich, distinctive extraction techniques, choosing the most suitable method for object recognition in industrial automation remains a challenge. Each technology differs in terms of computing cost, scale and lighting variation, and robustness compared to overall classification performance. This case study is to assess and compare the effectiveness of PCA, SIFT, and HOG functional descriptors for detecting 20 different objects recorded under different conditions. The aim is to determine which technology provides the best compromise between accuracy, efficiency, and simple implementation in a practical industrial environment.

3. Objectives

- To implement PCA, SIFT, and HOG feature extraction techniques on the COIL-20 dataset
- To train and evaluate a KNN classifier on features extracted by each method
- To compare the accuracy and performance of the techniques using standard evaluation metrics
- To identify the most effective feature extraction method for object recognition in industrial applications

4. Literature Review

Feature extraction is a fundamental step in image classification and object recognition systems. **Principal Component Analysis (PCA)** has long been used to reduce dimensionality while retaining the most informative components, thus improving computational efficiency without sacrificing performance [3], [13], [14]. PCA is particularly effective when the dataset is well-structured and the features are correlated [3].

Scale-Invariant Feature Transform (SIFT), introduced by Lowe [1], is a robust local feature descriptor known for its scale and rotation invariance. It has been widely adopted for tasks such as object recognition, matching, and retrieval. However, one of the challenges with SIFT is that it produces variable-length descriptors, requiring techniques like **Bag of Visual Words (BoVW)** [19] to create uniform feature representations.

Histogram of Oriented Gradients (HOG), developed by Dalal and Triggs [2], is effective for capturing edge-based and shape information. It is widely used in pedestrian detection and surveillance systems due to its high spatial locality and robustness to noise. HOG divides an image into cells and blocks to compute gradient orientation histograms, resulting in a fixed-length descriptor [2], [11].

Several studies [4], [5], [7] have shown that combining global and local descriptors can further improve classification performance. Additionally, KNN remains a simple yet effective classifier when paired with strong features [9], while more recent work explores integrating PCA and HOG into deep learning pipelines for hybrid approaches [6], [12].

5. **Contribution**

This study provides a practical comparative analysis of PCA, SIFT, and PIG for object recognition in industrial environments. Using the same classifier (KNN) and consistent data records (COIL-20) ensures a fair assessment of the effectiveness of individual methods. We also implement SIFT in relation to visual word bags to tackle variable length descriptor tasks. This study not only shows the accuracy of classification, but also shows the relative complexity of arithmetic and implementation complexity, providing developers and researchers who select features suitable for real-world industrial applications.

Methods

1. **Data Collection**

The dataset used in this study is the **COIL-20 (Columbia Object Image Library)**, which contains grayscale images of 20 different objects captured from varying angles. Each object has 72 images taken at 5-degree intervals, resulting in a total of 1,440 images. The objects are placed against a black background and are uniformly illuminated, making it suitable for evaluating feature extraction methods in a controlled setting. All images were resized to 128×128 pixels and converted to grayscale for consistency across methods.

For the purpose of this case study, the dataset was organized into 20 subfolders, each representing a unique object class. Labels were assigned based on folder names, and the images were read using OpenCV. Figure 1 shows sample images from the COIL-20 dataset. After loading and preprocessing, the data was split into training and testing sets using an 80:20 ratio, with stratified sampling to maintain class balance. This setup ensures that the evaluation is robust and each class is equally represented.

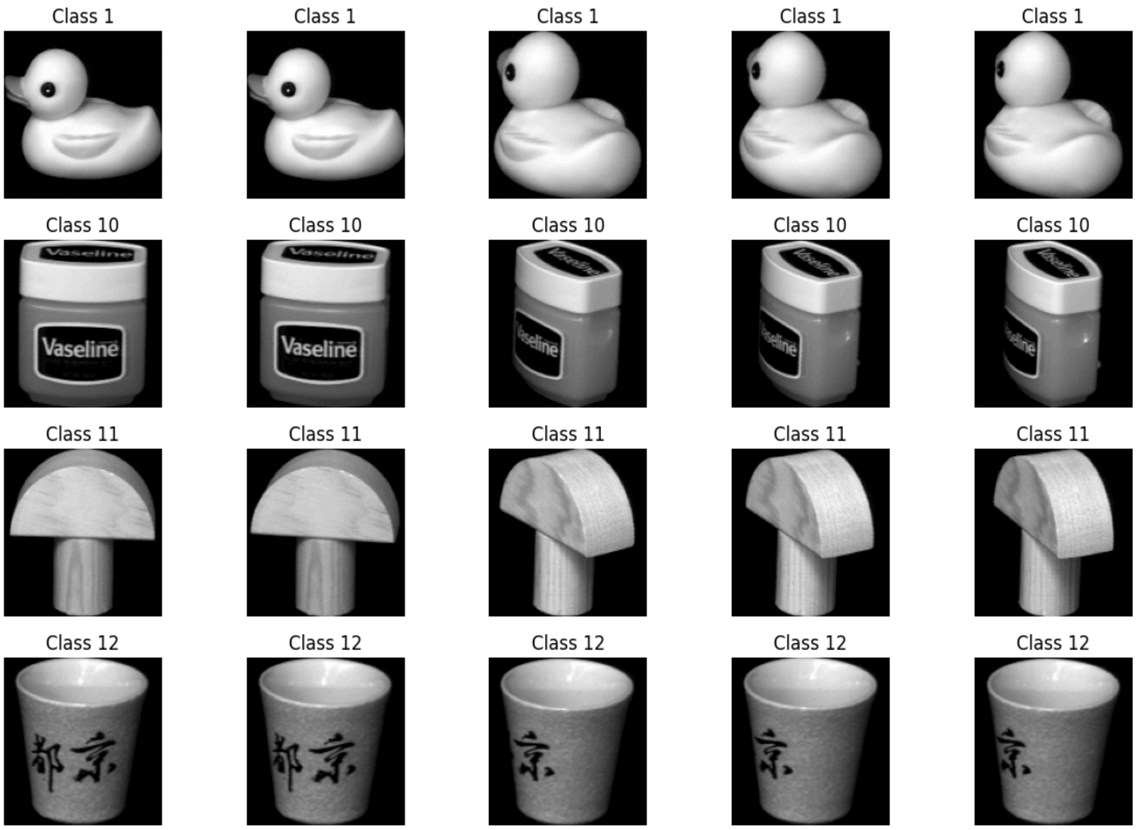

Figure 1: Sample grayscale images from four different object classes in the COIL-20 dataset. Each row represents one object category, with five images captured from different angles.

2. Algorithm/Model

This study compares three popular feature extraction techniques: **Principal Component Analysis (PCA)**, **Scale-Invariant Feature Transform (SIFT)**, and **Histogram of Oriented Gradients (HOG)**. Each technique was applied independently to the dataset, and a **K-Nearest Neighbours (KNN)** classifier was trained on the extracted features.

- **PCA**: The flattened images (128×128 = 16,384 features) were passed through a PCA transformation with 100 components. PCA reduced dimensionality by capturing the directions of maximum variance, resulting in a compact and noise-reduced representation. The transformed features were then used to train a KNN classifier.
- **SIFT + Bag of Visual Words (BoVW)**: SIFT was used to detect key points and extract 128-dimensional descriptors from each image. Since the number of descriptors per image varied, a **MiniBatch KMeans clustering** algorithm was used to build a vocabulary of 100 visual words. Each image was then

represented as a histogram showing the frequency of each visual word, forming a fixed-length vector suitable for KNN classification.

- **HOG**: HOG features were extracted using 9 orientation bins, 8×8 pixel cells, and 2×2 block normalization. This method captures object shape and gradient patterns, making it effective for structured object recognition. HOG features were directly used for training a KNN classifier.

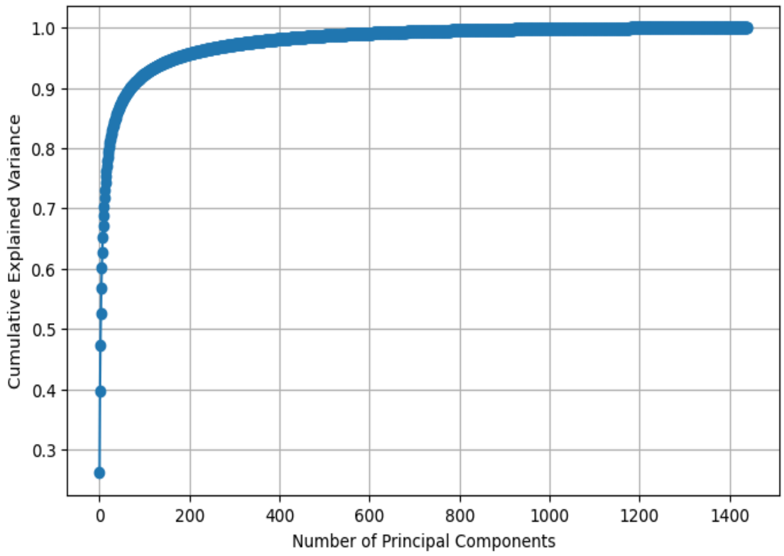

Figure 2: Variance Retained by PCA Components

3. Experimental Setup

All experiments were conducted using **Google Colab** with Python 3.10 and the following libraries: OpenCV, Scikit-learn, Scikit-image, NumPy, Matplotlib, and Seaborn. The feature extraction and classification pipelines were implemented and tested on an Intel Xeon virtual CPU with 12 GB RAM provided by Colab. Each model was trained on 80% of the dataset and tested on the remaining 20%.

To ensure fairness, all three techniques used the same train-test split, and the same classifier (KNN with k=3) was used across experiments. The number of PCA components and KMeans clusters was set to 100 based on standard practices for small to medium-sized image datasets.

4. **Evaluation Metrics**

To evaluate the performance of each feature extraction method, the following metrics were used:

- **Accuracy**: The percentage of correctly classified images.
- **Precision, Recall, and F1-score**: To evaluate the classifier's performance for each class.
- **Confusion Matrix**: To visualize the classification performance across all 20 classes.

Results

1. **Quantitative Results**

The performance of each feature extraction method — PCA, SIFT (with Bag of Visual Words), and HOG — was evaluated using classification accuracy, precision, recall, and F1-score. PCA + KNN achieved the highest accuracy of **98.96%**, followed by HOG at **98.26%**, while SIFT performed lower with **87.5%** accuracy. The results are summarized in **Table 1**, and visualized in **Figure 2** as a bar chart.

Classification reports were also generated for each method. PCA and HOG showed almost perfect precision and recall across most object classes, whereas SIFT displayed inconsistencies, particularly for objects with fewer key points or low-contrast regions. The confusion matrix in **Figure 3** (for PCA) demonstrates minimal class overlap, confirming the effectiveness of PCA-based feature representation.

These metrics highlight the importance of choosing feature extraction techniques based on the dataset and task complexity.

Feature Extraction Technique	**Accuracy (%)**
PCA	98.96
SIFT	87.50
HOG	98.26

Table 1: Accuracy Comparison Across PCA, SIFT, and HOG

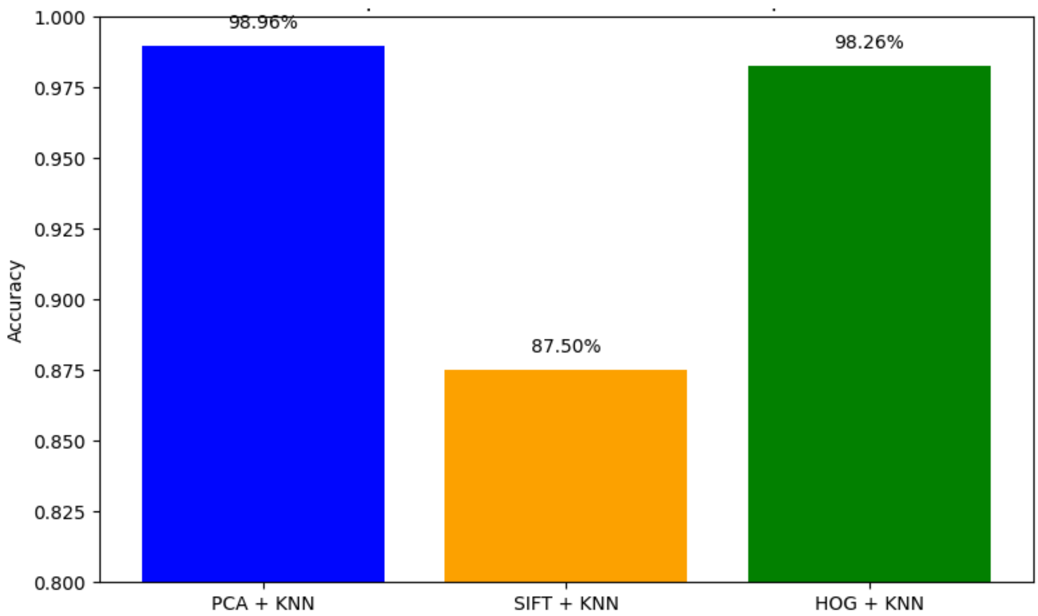

Figure 3: Bar Chart of Accuracies

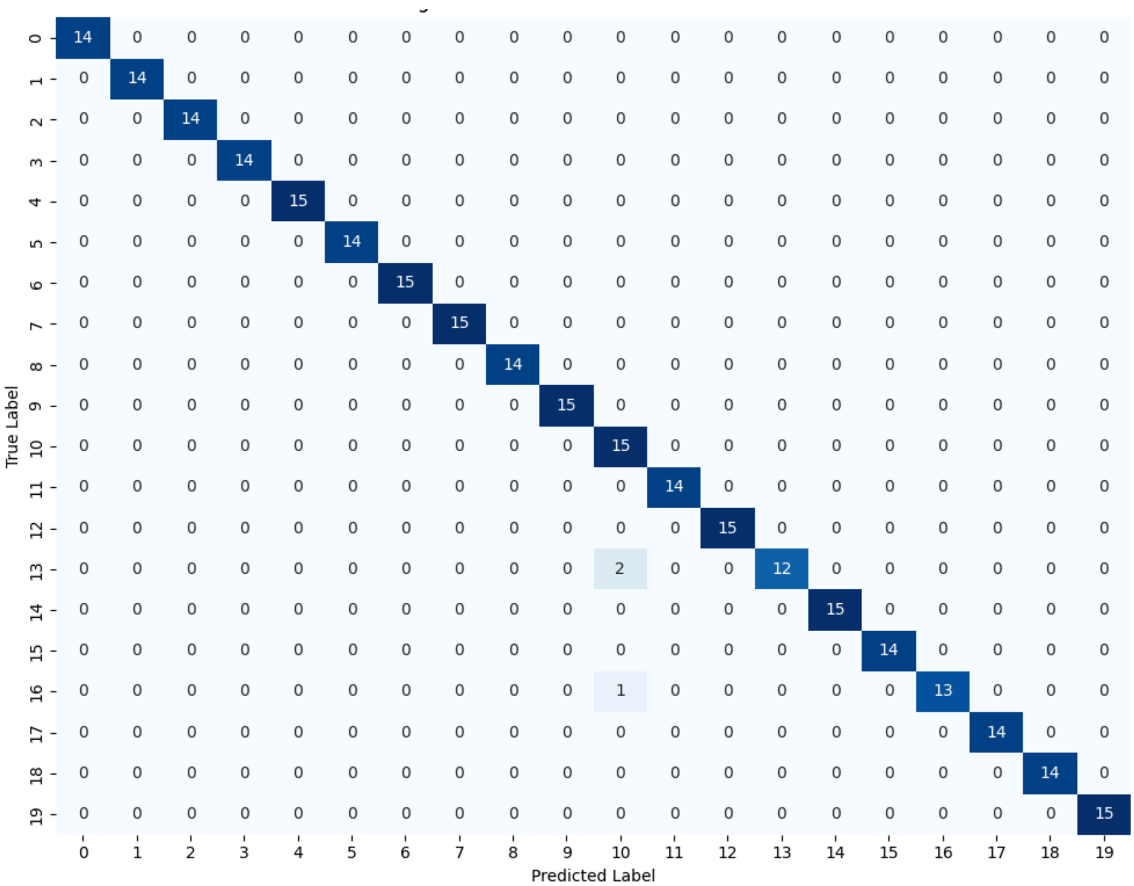

Figure 4: Confusion matrix for object classification using PCA-based features and KNN classifier

2. Qualitative Results

In addition to numerical metrics, visualizations were used to understand the feature extraction behaviour of SIFT and HOG. **Figure 4** illustrates SIFT keypoints detected on a sample image, where localized interest points are identified with scale and orientation information. These keypoints form the basis of the Bag of Visual Words histogram used for classification.

Figure 5 shows the HOG feature visualization for the same image. HOG focuses on capturing the distribution of gradients and edge directions, which represent object shape and structure effectively.

Both visualizations provide insight into how the techniques capture different aspects of the image: SIFT being more sparse and localized, and HOG being denser and more holistic in its representation. These qualitative observations align with the quantitative results, where HOG performed closer to PCA, and SIFT slightly lagged behind.

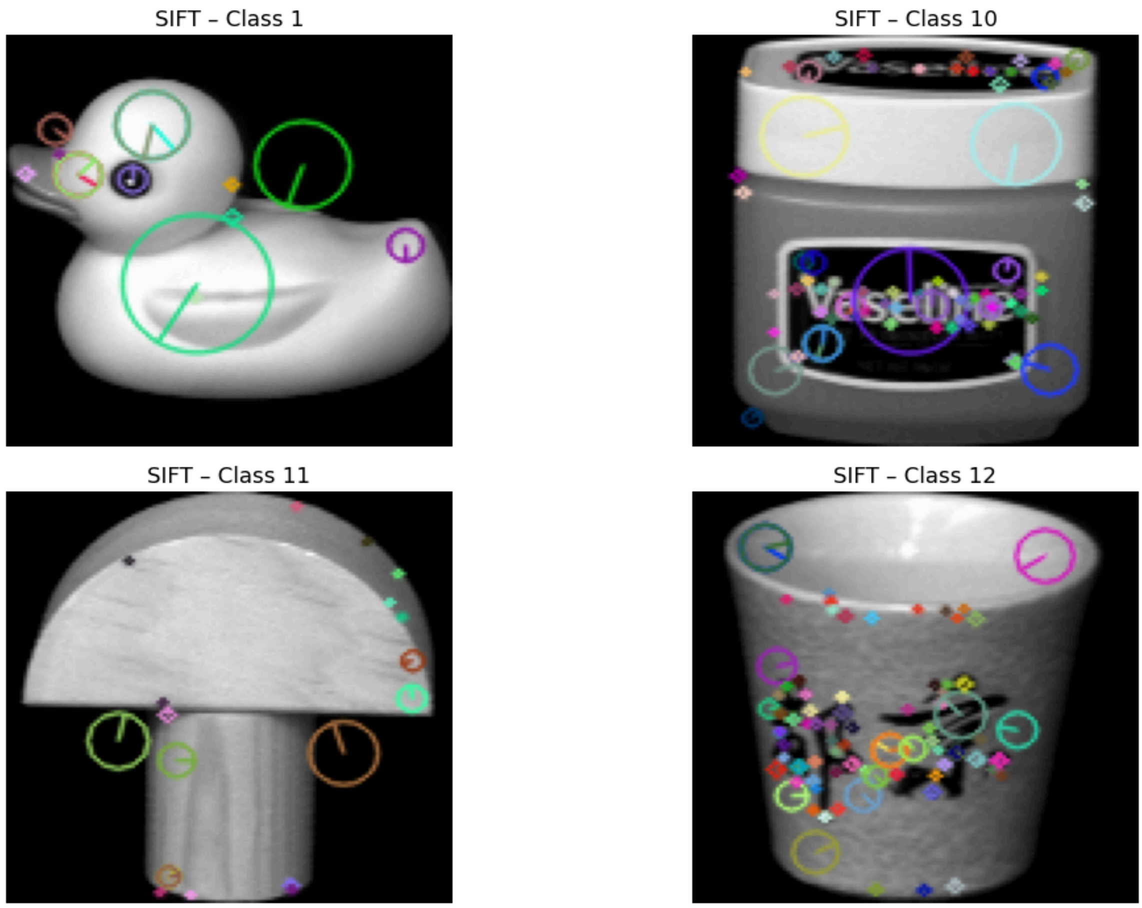

Figure 5: SIFT keypoints visualized on four different object classes from the COIL-20 dataset.

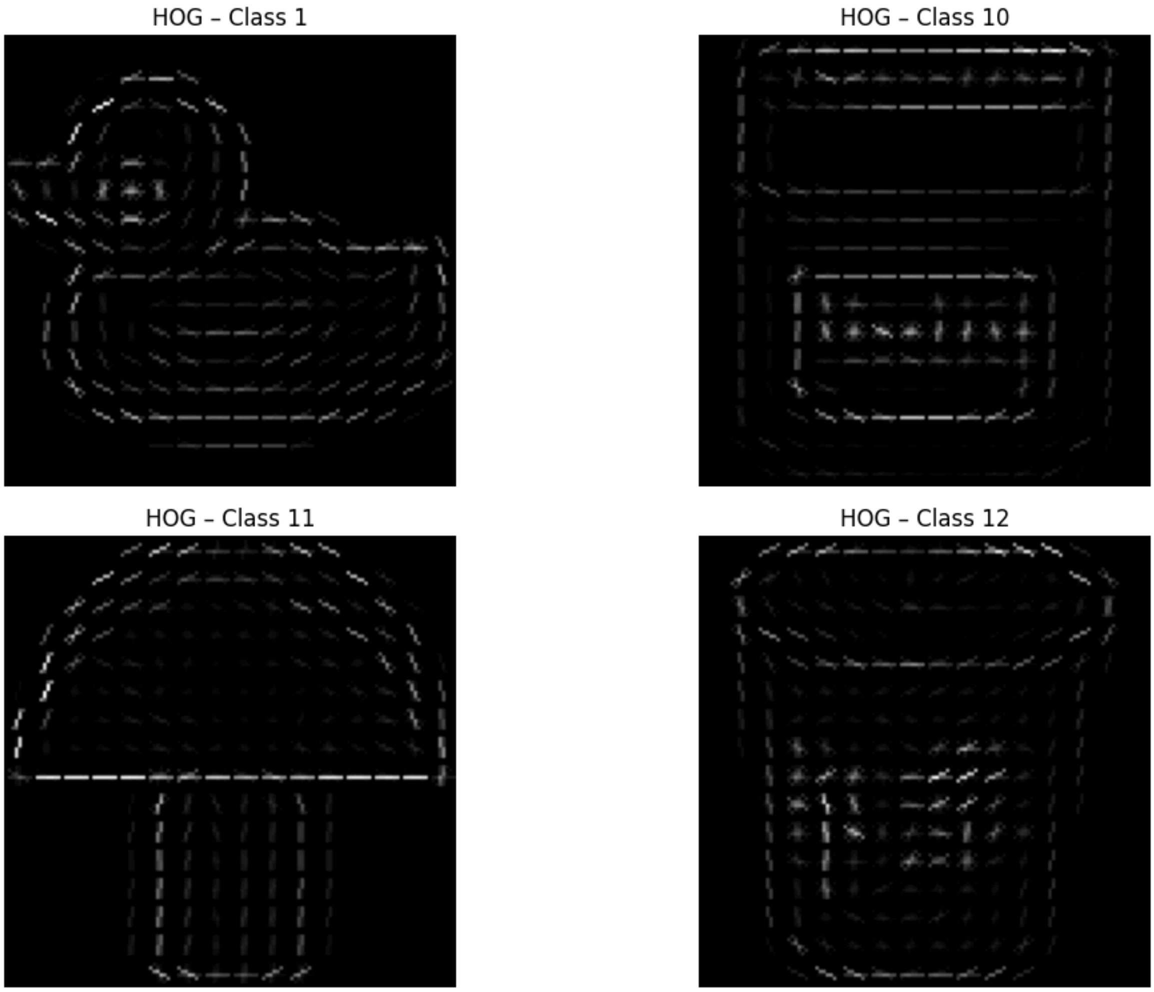

Figure 6: HOG feature visualization on four different object classes from the COIL-20 dataset.

3. Comparison with Baselines

Several studies have used feature extraction techniques such as SIFT and HOG in object recognition tasks, and PCA is frequently used for dimensionality reduction in image classification pipelines. In our experiment, PCA outperformed both SIFT and HOG in terms of accuracy, despite being the simplest to implement.

These findings suggest that global statistical techniques like PCA can sometimes outperform handcrafted local features when the dataset is clean and consistently structured, as in the COIL-20 dataset. In contrast, SIFT and HOG are better suited for real-world scenarios where object deformation, lighting, and occlusion play a significant role.

While this study used KNN as a common classifier, future comparisons with advanced deep learning methods could further benchmark the true effectiveness of these classical techniques.

Pitfall

One major pitfall in image classification experiments is the overreliance on accuracy as the sole performance metric. Accuracy can be misleading in imbalanced datasets or multi-class scenarios where precision and recall may vary significantly across classes. Furthermore, SIFT's variable-length descriptors required conversion using a BoVW model, which itself introduces quantization error and affects final performance.

Solution

To address this, we employed multiple evaluation metrics, including precision, recall, F1-score, and confusion matrices, to assess per-class performance more comprehensively. We also ensured stratified train-test splitting to maintain class balance. Additionally, for SIFT, we used **MiniBatch KMeans** to reduce quantization loss in the Bag of Visual Words representation.

Discussion

1. Interpretation of Results

The experimental results indicate that PCA-based features outperform both SIFT and HOG in terms of classification accuracy on the COIL-20 dataset when used with a K-Nearest Neighbours classifier. PCA achieved an impressive accuracy of **98.96%**, followed closely by HOG at **98.26%**, while SIFT lagged behind at **87.5%**. These findings suggest that PCA, despite being a linear and unsupervised dimensionality reduction technique, effectively captures the most relevant global patterns in the images. HOG also demonstrated strong performance, likely due to its ability to encode structural gradients and object shapes effectively. The lower performance of SIFT may be attributed to the quantization error introduced by the Bag of Visual Words model and the loss of fine detail in the descriptor aggregation process.

These findings are significant for image processing and object recognition tasks, as they show that simpler, global statistical methods like PCA can be more effective than complex local descriptors under controlled imaging conditions. Additionally, the results validate the importance of choosing the right feature extraction technique based on the nature of the dataset — with PCA being well-suited for uniformly captured object datasets such as COIL-20. The consistent classification performance across most classes further reinforces the robustness of PCA and HOG as feature representation techniques in structured environments.

2. **Limitations**

While the results are promising, this study has a few limitations. The dataset used (COIL-20) consists of clean, centred, and uniformly captured grayscale images with minimal background noise or occlusion, which may not reflect real-world complexity. The evaluation was also limited to a single classifier (KNN) and a fixed number of features or clusters, which may not generalize across all tasks. SIFT's performance may have improved with better tuning of the vocabulary size or by combining it with more sophisticated classifiers. Future work could explore performance under more realistic datasets, evaluate additional classifiers like SVM or CNNs, and experiment with hybrid feature extraction strategies to combine global and local features more effectively.

3. **Practical Implications**

The results of this study have meaningful practical implications for industrial automation and real-time object recognition systems. PCA's high performance and computational efficiency make it a strong candidate for embedded systems where resources are limited. Similarly, HOG's edge-based feature representation makes it ideal for shape recognition in quality control or inspection lines. The insights from this work can help practitioners choose the appropriate feature extraction method based on task-specific constraints such as lighting, background variability, or computational capacity. In scenarios where high-speed recognition of static, structured objects is required, PCA or HOG can be reliably deployed. Furthermore, the simplicity of implementation and high interpretability of these techniques make them attractive choices for early-stage development and prototyping in vision-based automation systems.

Conclusion

1. **Summary**

This study presented a comparative evaluation of three feature extraction techniques — Principal Component Analysis (PCA), Scale-Invariant Feature Transform (SIFT), and Histogram of Oriented Gradients (HOG) — for object recognition using the COIL-20 dataset. PCA achieved the highest classification accuracy of **98.96%**, followed closely by HOG with **98.26%**, while SIFT recorded a lower accuracy of **87.5%**. The results demonstrate that global statistical features such as those extracted by PCA can be highly effective for structured object datasets, outperforming more complex local descriptors like SIFT. Additionally, HOG proved to be a robust alternative, especially for capturing shape-based features.

2. Future Work

Future research can explore the integration of these classical techniques with modern deep learning models to combine interpretability with performance. Expanding the experiments to include datasets with cluttered backgrounds, varying lighting conditions, or partial occlusion would provide a more comprehensive evaluation of robustness. Evaluating alternative classifiers such as SVMs, Random Forests, or neural networks could also yield deeper insights into feature-classifier combinations. Additionally, hybrid approaches that merge global and local descriptors may offer improved performance in more diverse, real-world applications.

References

[1] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[2] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2005, pp. 886–893.

[3] I. Jolliffe, *Principal Component Analysis*, 2nd ed. New York: Springer, 2002.

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.

[5] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. Wiley-Interscience, 2001.

[6] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[8] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

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

[10] B. K. P. Horn, "Understanding image gradients," *Artificial Intelligence Laboratory Memo*, MIT, 1986.

[11] G. Bradski and A. Kaehler, *Learning OpenCV: Computer Vision with the OpenCV Library*, O'Reilly Media, 2008.

[12] G. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[13] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[14] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed., Springer, 2009.

[15] J. C. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," *Advances in Large Margin Classifiers*, pp. 61–74, 1999.

[16] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed., Academic Press, 1990.

[17] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2014.

[18] M. Everingham et al., "The Pascal Visual Object Classes (VOC) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010.

[19] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE CVPR*, 2006, pp. 2169–2178.

[20] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.