**IAES Journal of Electrical and Electronics Engineering**
Vol. 1, No. 1, June 2014, pp. 10~14
ISSN: 2338-7432, DOI: 10.11591/jeee.v1i1.pp10-14 

# Object recognition and classification: from handcrafted features to learning-based approaches

**Anto, Tole Sutikno**
Faculty of Industrial Technology, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

**Article Info**

***Article history:***

Received Feb 13, 2014
Revised Apr 19, 2014
Accepted May 2, 2014

***Keywords:***

Computer vision
Feature engineering
Handcrafted features
Image classification
Machine learning
Object recognition
SVM

**ABSTRACT**

Object recognition and classification are fundamental tasks in computer vision that aim to identify and categorize objects based on visual information. Over the years, these tasks have evolved from traditional approaches based on geometric representations and handcrafted features toward machine learning and learning-based recognition systems. This paper presents a critical review of the development of object recognition, focusing on feature engineering, support vector machines (SVM), and classifier-based recognition. Particular attention is given to handcrafted feature representations, including Haar-like features, scale-invariant feature transform (SIFT), histograms of oriented gradients (HOG), and Bag-of-Visual-Words (BoVW), together with classification algorithms such as SVM, AdaBoost, and decision trees. The review also discusses the emergence of data-driven approaches, particularly early convolutional neural networks, as an important transition in visual recognition. The analysis highlights the strengths and limitations of conventional recognition pipelines and emphasizes the importance of feature representation, classifier selection, and learning capability in developing robust object recognition systems.

*This is an open access article under the CC BY-SA license.*

## 1. INTRODUCTION

Object recognition is a fundamental research area in computer vision that enables machines to identify and distinguish objects from digital images. It involves analyzing visual characteristics, extracting meaningful information, and assigning objects to predefined categories. Its applications include industrial inspection, autonomous systems, surveillance, medical imaging, robotics, and intelligent transportation.

Early object recognition systems primarily relied on geometric models, template matching, and manually designed visual representations. These approaches attempted to identify objects by comparing image structures, shapes, edges, and other predefined characteristics. However, variations in illumination, scale, orientation, occlusion, and background complexity significantly limited their generalization capabilities.

The introduction of machine learning substantially changed the development of object recognition systems. Instead of relying exclusively on predefined decision rules, researchers began combining engineered visual features with statistical learning algorithms such as support vector machines (SVM), AdaBoost, and decision trees. This development established a more systematic recognition framework in which feature extraction and classifier training jointly determine recognition performance.

This paper critically reviews the evolution of object recognition and classification, emphasizing feature engineering, conventional classifiers, and emerging learning-based approaches. The discussion examines their fundamental principles, advantages, limitations, and contributions to the development of computer vision. The review considers research published up to 2013, including the early emergence of deep learning as a transition toward automatically learned visual representations.

## 2. EVOLUTION OF OBJECT RECOGNITION

The development of object recognition can be understood through the transition from geometric and appearance-based methods toward feature-driven and learning-based systems. Early approaches emphasized explicit object representations, including edges, contours, corners, and geometric relationships. Although these methods were effective in controlled environments, they struggled with complex scenes and significant variations in object appearance.

A major development occurred with the increasing use of local visual features and statistical pattern recognition. Viola and Jones [1] demonstrated the effectiveness of combining Haar-like features, integral images, AdaBoost, and cascaded classification for rapid face detection. Their approach established an influential framework for object detection by combining efficient feature computation with machine learning-based classification.

Subsequent research introduced more sophisticated feature representations that improved robustness against image transformations. scale-invariant feature transform (SIFT), developed by Lowe [2], provided distinctive local descriptors that were relatively invariant to scale and rotation, while histograms of oriented gradients (HOG), introduced by Dalal and Triggs [3], demonstrated the effectiveness of gradient-based representations for human detection. These developments contributed to the widespread adoption of feature-based recognition pipelines.

By the early 2010s, object recognition increasingly incorporated more complex learning mechanisms and structured object representations. Deformable part models (DPM), introduced by Felzenszwalb *et al.* [4], combined local appearance features with spatial relationships between object parts. This development demonstrated that effective recognition required not only discriminative features but also appropriate representations of object structure.

## 3. FEATURE ENGINEERING FOR OBJECT RECOGNITION

Feature engineering plays a central role in conventional object recognition because it determines how visual information is transformed into representations suitable for classification. The process involves selecting, designing, and extracting characteristics that distinguish one object category from another. The quality of these features directly influences the ability of a classifier to recognize objects under different visual conditions.

Several handcrafted feature representations have become influential in object recognition. Haar-like features capture differences in image intensity across rectangular regions and are particularly effective when combined with integral images and AdaBoost. SIFT describes distinctive local image structures through gradient information, while HOG represents object appearance through distributions of local gradient orientations.

Another important approach is the Bag-of-Visual-Words (BoVW) model, which represents images through the occurrence of visual patterns derived from local descriptors. Sivic and Zisserman [5] introduced the visual vocabulary concept for matching objects and scenes in videos, while Csurka *et al.* [6] demonstrated its application to visual categorization. This representation enables local descriptors to be aggregated into fixed-length feature vectors that can be processed by conventional classifiers.

Critical evaluation: Handcrafted features provide interpretable and computationally manageable representations when their characteristics match the recognition problem. Nevertheless, their effectiveness depends heavily on feature design, parameter selection, and the visual properties of the target objects. They may fail to capture complex semantic information because their representations are generally determined by predefined mathematical and structural assumptions.

## 4. CLASSIFIER-BASED OBJECT RECOGNITION

Classifier-based recognition combines extracted visual features with machine learning algorithms to determine the category of an object. Unlike simple feature matching, this approach learns decision boundaries or classification rules from labeled training samples. The choice of classifier is therefore an important factor influencing recognition accuracy, computational efficiency, and generalization.

### 4.1. Support vector machines

SVM is a supervised learning algorithm that identifies discriminative decision boundaries between object categories. Cortes and Vapnik [7] established the theoretical foundation of support vector networks, while subsequent research demonstrated their effectiveness in visual classification. By using different kernel functions, SVM can handle nonlinear relationships between feature representations and object categories.

SVM has been widely adopted in object recognition because it performs effectively with high-dimensional feature vectors and relatively limited training datasets. It is particularly suitable for classification

systems using SIFT, HOG, and BoVW representations. However, its performance depends on kernel selection, parameter configuration, feature quality, and training data characteristics.

### 4.2. AdaBoost and cascaded classification

AdaBoost is an ensemble learning approach that combines multiple weak classifiers to construct a stronger classification model. Freund and Schapire [8] established its theoretical foundation, while Viola and Jones [1] demonstrated its practical effectiveness for rapid face detection. Their cascaded architecture progressively rejects non-object regions, allowing computational resources to concentrate on promising detection candidates.

The main advantage of AdaBoost is its ability to combine simple features and weak learners into an effective recognition system. It is particularly useful when computational efficiency and rapid detection are important requirements. Nevertheless, its performance can be affected by training sample quality, feature selection, and changes in object appearance.

### 4.3. Other classification approaches

Several other classification algorithms have contributed to conventional object recognition systems. K-nearest neighbors (KNN) classifies objects based on similarity to previously labeled samples, while decision trees use hierarchical decision rules to distinguish object categories. Random forests further extend decision-tree approaches by combining multiple trees to improve classification stability.

These classifiers offer different balances between computational requirements, model complexity, and classification performance. KNN is relatively simple but can become expensive during inference when training datasets are large. Decision trees and random forests provide flexible classification mechanisms, although their effectiveness depends on feature representation and appropriate model configuration.

## 5. FROM HANDCRAFTED FEATURES TO LEARNING-BASED APPROACHES

The transition from handcrafted features to learning-based approaches represents an important change in the design of object recognition systems. Conventional pipelines separate feature engineering and classifier training into distinct stages, requiring researchers to determine which visual characteristics are most relevant. Learning-based approaches increasingly attempt to discover useful representations directly from training data.

Convolutional neural networks (CNNs) represent an important development in this direction. Earlier work by LeCun *et al.* [9] demonstrated the effectiveness of convolutional neural networks for handwritten character recognition through automatically learned hierarchical representations. Their work established an important foundation for subsequent applications of neural networks to visual recognition.

A major milestone occurred in 2012 when Krizhevsky *et al.* [10] demonstrated the effectiveness of deep convolutional neural networks in large-scale image classification through AlexNet. Their results highlighted the potential of learning hierarchical visual features directly from image data rather than depending entirely on manually engineered descriptors. This achievement marked a significant shift in research attention toward data-driven visual representation and large-scale learning.

Critical evaluation: Learning-based approaches reduce dependence on manual feature design and can capture more complex visual patterns than conventional handcrafted representations. However, their effectiveness depends on training data availability, computational resources, model configuration, and optimization procedures. By 2013, these approaches had demonstrated substantial potential, although conventional feature-based methods remained relevant for applications with limited data and computational resources [11].

## 6. COMPARATIVE ANALYSIS AND DISCUSSION

The reviewed approaches differ in their feature representation strategies, learning mechanisms, computational requirements, and generalization capabilities. Handcrafted features provide predefined visual representations, whereas learning-based approaches increasingly derive representations from training data. Classifier-based recognition connects these representations to category prediction through supervised or ensemble learning mechanisms.

Table 1 summarizes the main characteristics, advantages, and limitations of representative object recognition approaches.

Table 1. Comparative analysis of object recognition and classification approaches

| Approach | Main advantage | Main limitation |
|---|---|---|
| Haar-like features | Efficient feature computation | Limited representation of complex structures |
| SIFT | Distinctive local feature representation | Computationally intensive |
| HOG | Effective shape and gradient representation | Sensitive to significant geometric variations |
| Bag-of-visual-words | Flexible aggregation of local features | Limited spatial structure representation |
| SVM | Effective high-dimensional classification | Sensitive to kernel and parameter selection |
| AdaBoost | Efficient ensemble and cascaded classification | Dependent on feature and training sample quality |
| KNN | Simple similarity-based classification | Expensive inference for large datasets |
| Decision trees/Random forests | Flexible and interpretable classification | Performance depends on feature quality |
| CNN | Automatically learns hierarchical features | Requires substantial training data and computation |

The comparison demonstrates that recognition performance cannot be attributed to the classifier alone. Feature representation, training data, object variability, and classifier configuration jointly influence the final outcome. Consequently, selecting an appropriate combination of feature extraction and classification techniques is essential for developing reliable recognition systems.

An important observation is that the transition toward learning-based recognition does not simply represent the replacement of one classifier by another. It reflects a broader change in how visual representations are developed and optimized. The ability to learn discriminative features directly from data offers an alternative to conventional feature engineering, although computational and data requirements remain important considerations.

## 7. RESEARCH CHALLENGES AND FUTURE DIRECTIONS

Despite substantial progress up to 2013, several challenges remained in object recognition and classification. Variations in illumination, viewpoint, scale, occlusion, and background complexity continued to affect recognition reliability. These challenges indicated the need for more robust visual representations and classification strategies.

Another important challenge concerns the relationship between feature quality and classifier performance. Combining multiple handcrafted features may improve visual representation, but it can also introduce redundancy and computational complexity. Future research should therefore emphasize more effective feature integration, representation learning, and classifier optimization.

The emergence of deep learning also introduced new research opportunities for automatically learning hierarchical visual representations. Further improvements in training strategies, computational efficiency, and generalization were important directions for subsequent research. The integration of feature engineering insights with data-driven learning also offered opportunities for developing more effective recognition systems.

## 8. CONCLUSION

This paper reviewed the development of object recognition and classification, focusing on feature engineering, SVM, classifier-based recognition, and the transition toward learning-based approaches. Handcrafted features such as Haar-like features, SIFT, and HOG established important foundations for conventional recognition systems. Their integration with SVM, AdaBoost, and other classifiers significantly improved the effectiveness of automated object recognition. The emergence of convolutional neural networks represented a major transition toward automatically learned visual representations, particularly following the achievements demonstrated by AlexNet in 2012. Nevertheless, conventional feature-based approaches remained valuable because of their interpretability, flexibility, and suitability for resource-constrained applications. Ultimately, effective object recognition depends on the appropriate integration of visual representation, learning mechanisms, classification strategies, and application requirements.

## REFERENCES

[1] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition*, IEEE Comput. Soc, 2001, pp. I-511-I–518. doi: 10.1109/cvpr.2001.990517.

[2] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004, doi: 10.1023/B:VISI.0000029664.99615.94.

[3] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, IEEE, 2005, pp. 886–893. doi: 10.1109/CVPR.2005.177.
[4] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627–1645, Sep. 2010, doi: 10.1109/TPAMI.2009.167.
[5] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *Proceedings ninth IEEE international conference on computer vision*, IEEE, 2003, pp. 1470–1477. doi: 10.1109/iccv.2003.1238663.
[6] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Proceedings of the Workshop on Statistical Learning in Computer Vision, European Conference on Computer Vision (ECCV)*, 2004, pp. 1–22.
[7] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/bf00994018.
[8] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, Aug. 1997, doi: 10.1006/jcss.1997.1504.
[9] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998, doi: 10.1109/5.726791.
[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
[11] A. Andreopoulos and J. K. Tsotsos, "50 Years of object recognition: Directions forward," *Computer Vision and Image Understanding*, vol. 117, no. 8, pp. 827–891, Aug. 2013, doi: 10.1016/j.cviu.2013.04.005.