IAES Journal of Electrical and Electronics Engineering
Vol. 1, No. 1, June 2014, pp. 5~9
ISSN: 2338-7432, DOI: 10.11591/jeee.v1i1.pp5-9

Feature extraction and feature selection in image classification: a critical review

Romi Hadini, Sunardin
Faculty of Industrial Technology, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

Article Info

Article history:

Received Dec 26, 2013
Revised Mar 20, 2014
Accepted Apr 15, 2014

Keywords:

Dimensionality reduction
Feature extraction
Feature selection
HOG
Image classification
Local descriptors
SIFT

ABSTRACT

Feature extraction and feature selection are fundamental processes in image classification because they determine how visual information is represented and which characteristics are used by a classifier. Conventional image classification systems rely on handcrafted features to describe local structures, object shapes, textures, and visual patterns. This paper presents a critical review of major feature representation techniques, focusing on scale-invariant feature transform (SIFT), speeded-up robust features (SURF), histograms of oriented gradients (HOG), and other local descriptors. It also examines dimensionality reduction and feature selection methods for improving classification efficiency and predictive performance. The analysis highlights the complementary roles of feature extraction, dimensionality reduction, and feature selection, while identifying their limitations in terms of computational complexity, feature redundancy, sensitivity to image variations, and information loss. The review emphasizes that effective image classification requires not only discriminative feature representations but also appropriate strategies for selecting and reducing feature dimensions.

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

1. INTRODUCTION

Image classification is an important task in computer vision and pattern recognition that aims to assign images to predefined categories based on their visual characteristics. The performance of an image classification system depends not only on the classifier but also on the quality and discriminative capability of the extracted features. Consequently, feature extraction and feature selection play essential roles in developing accurate and computationally efficient classification systems.

Traditional image classification approaches commonly use handcrafted features to represent image characteristics, including color, texture, shape, and local structures. Several influential techniques have been developed to capture these characteristics, including scale-invariant feature transform (SIFT), speeded-up robust features (SURF) and histograms of oriented gradients (HOG). Although these methods provide effective visual representations, they may generate high-dimensional feature vectors, redundant information, and computational challenges.

This paper critically reviews feature extraction and feature selection techniques for image classification, emphasizing local descriptors, dimensionality reduction, and feature selection. The discussion examines their fundamental principles, representative methods, advantages, and limitations. The objective is to provide an integrated understanding of how feature representation and feature optimization influence image classification performance, based exclusively on literature published up to 2013.

2. FEATURE EXTRACTION TECHNIQUES

2.1. Scale-invariant feature transform

SIFT is a widely recognized local feature extraction technique designed to identify distinctive image structures that remain relatively stable under changes in scale and rotation. Introduced by Lowe [1], SIFT

detects keypoints using scale-space analysis and describes their local neighborhoods through gradient orientation distributions. Its representation provides robustness against moderate changes in illumination, viewpoint, image scale, and rotation.

SIFT has been extensively applied in object recognition, image matching, image retrieval, and visual classification. Its main advantage is its ability to identify distinctive local structures even when images contain partial occlusion or substantial background variations. However, SIFT generates relatively high-dimensional descriptors and may require considerable computational resources when processing large image collections.

2.2. Speeded-up robust features

SURF was developed to provide an efficient alternative to conventional local feature descriptors while maintaining robustness to image transformations. Bay *et al.* [2] introduced SURF using approximations of second-order Gaussian derivatives and integral images to accelerate keypoint detection and descriptor computation. Its design provides a balance between computational efficiency and local feature distinctiveness.

Compared with SIFT, SURF generally offers faster feature extraction and is attractive for applications requiring relatively efficient image processing. Nevertheless, its performance may decrease under substantial viewpoint changes, complex transformations, or image conditions that affect local feature detection. Therefore, SURF should not automatically be considered superior to SIFT, because their relative performance depends on the image characteristics and classification requirements.

2.3. Histograms of oriented gradients

HOG is a feature extraction technique that represents local image appearance through distributions of gradient orientations. Dalal and Triggs [3] demonstrated its effectiveness for human detection by combining local gradient histograms with spatial cells and block normalization. This representation captures object shape and edge structures while providing some robustness to local illumination variations.

HOG is particularly effective for object recognition and classification tasks in which shape and contour information are important. Unlike sparse keypoint-based approaches, HOG typically computes descriptors over regularly distributed image regions, providing a more systematic representation of the overall object structure. However, HOG can produce high-dimensional feature vectors and may be sensitive to significant geometric transformations, object deformation, and variations in spatial alignment.

2.4. Local descriptors and other feature representations

Local descriptors provide a means of representing distinctive image regions rather than relying exclusively on global image characteristics. In addition to SIFT and SURF, several descriptors have been developed to capture local texture, intensity, and structural patterns. Examples include local binary Patterns (LBP), Gabor-based features, and gradient-based descriptors.

LBP represents local texture through relationships between neighboring pixel intensities and has been widely applied in texture classification and face analysis [4]. Gabor features capture spatial frequency and orientation information, making them useful for describing texture patterns at different scales. The performance of these descriptors depends on the type of visual information that is most relevant to the classification problem.

An important development in local feature representation is the bag-of-visual-words (BoVW) model, which converts multiple local descriptors into a fixed-length image representation. This approach groups local descriptors into visual vocabulary clusters and represents each image through the distribution of visual words [5]. Although BoVW provides a practical framework for image classification, it may lose spatial relationships between local features and introduce additional computational requirements during vocabulary construction.

3. DIMENSIONALITY REDUCTION

Dimensionality reduction addresses the problem of high-dimensional feature representations by transforming the original feature space into a smaller and more manageable representation. This process is important because local descriptors and combined feature representations may contain substantial redundancy and correlated information. Reducing feature dimensions can decrease computational costs, simplify classifier training, and potentially improve generalization.

Principal component analysis (PCA) is one of the most widely used dimensionality reduction techniques. It transforms the original feature space into a set of orthogonal components that preserve as much of the data variance as possible [6]. PCA has also been applied to local image descriptors, including PCA-SIFT, to obtain more compact representations [7].

Linear discriminant analysis (LDA) provides another approach by identifying feature projections that maximize the separation between predefined classes. Unlike PCA, which focuses on preserving overall data variance, LDA explicitly considers class information during dimensionality reduction. However, its effectiveness depends on the availability and representativeness of labeled training samples.

Other approaches include independent component analysis (ICA), random projections, and manifold-based methods such as Isomap and locally linear embedding. These methods offer different strategies for representing complex data structures and reducing feature dimensions. Nevertheless, dimensionality reduction may discard discriminative information, particularly when the retained components do not adequately represent differences between image classes.

4. FEATURE SELECTION TECHNIQUES

Feature selection aims to identify the most relevant features from an existing feature set while removing irrelevant, redundant, or noisy information. Unlike dimensionality reduction, which transforms the original features into a new representation, feature selection retains a subset of the original features. This distinction is important because selected features generally preserve their original interpretation and may improve the interpretability of classification systems.

Feature selection methods are commonly categorized into filter, wrapper, and embedded approaches [8]. Filter methods evaluate feature relevance using statistical or information-theoretic criteria without directly involving a particular classifier. Examples include information gain, correlation-based selection, mutual information, and statistical ranking.

Wrapper methods evaluate feature subsets through the performance of a selected classification algorithm. They can identify feature combinations that are particularly useful for a specific classifier, but their repeated model evaluation may introduce substantial computational costs. Embedded methods perform feature selection during classifier training, with representative examples including decision trees and regularized learning algorithms.

Mutual information-based selection is particularly relevant when image features contain complex relationships with class labels. Peng *et al.* [9] proposed the minimum-redundancy maximum-relevance approach, which balances feature relevance and redundancy. This principle is valuable for image classification because combining multiple descriptors may produce overlapping information that does not necessarily improve classification accuracy.

5. COMPARATIVE ANALYSIS AND DISCUSSION

Feature extraction, dimensionality reduction, and feature selection address different but interconnected challenges in image classification. Feature extraction determines how visual characteristics are represented, dimensionality reduction creates a more compact representation, and feature selection identifies the most informative original features. Their appropriate integration can improve classification efficiency while reducing unnecessary computational complexity [10]. Table 1 summarizes the main characteristics, advantages, and limitations of the reviewed techniques.

Table 1. Comparative analysis of feature extraction and feature optimization techniques

Technique	Main advantage	Main limitation
SIFT	Robust local features under scale and rotation changes	High computational cost and descriptor dimensionality
SURF	Faster local feature extraction	Sensitivity to certain complex transformations
HOG	Effective representation of shape and gradient information	High dimensionality and spatial alignment sensitivity
LBP	Simple and effective local texture representation	Limited representation of global structures
Gabor features	Captures texture at multiple scales and orientations	Computational complexity and parameter sensitivity
BoVW	Converts local descriptors into fixed-length representations	Loss of spatial information
PCA	Compact representation and redundancy reduction	May discard discriminative information
LDA	Improves class separation	Requires labeled data and depends on class distributions
Feature selection	Retains relevant original features and reduces redundancy	Sensitive to selection criteria and classifier characteristics

The comparison indicates that no single feature representation is universally suitable for all image classification tasks. SIFT and SURF are particularly useful when distinctive local structures are important, whereas HOG is more appropriate when object shape and gradient distributions provide discriminative information. LBP and Gabor features offer complementary capabilities for texture-oriented classification problems [11].

Dimensionality reduction and feature selection also have different implications for classification performance. PCA can provide compact representations, but preserving high data variance does not necessarily guarantee the preservation of class-discriminative information. Feature selection, in contrast, can retain meaningful original descriptors, although its effectiveness depends on the relevance criteria and interactions among selected features [12].

An important research consideration is the relationship between feature representation and classifier design. A feature set that performs effectively with one classifier may not provide similar results with another classifier. Therefore, feature evaluation should consider classification accuracy, computational complexity, robustness, feature redundancy, and generalization across different datasets.

6. RESEARCH CHALLENGES AND FUTURE DIRECTIONS

Despite considerable progress before 2014, several challenges remained in developing effective feature-based image classification systems. First, handcrafted descriptors may not adequately represent complex visual variations, particularly when images contain substantial changes in illumination, scale, orientation, and object appearance. Second, combining multiple descriptors can improve feature diversity but may also introduce redundancy and increase computational requirements.

Another important challenge concerns the selection of optimal feature dimensions and subsets. Excessive dimensionality reduction may remove useful information, whereas insufficient reduction may increase computational complexity and the risk of overfitting. Future research should therefore emphasize adaptive feature optimization that considers both class-discriminative information and computational efficiency.

Hybrid feature representation also offers an important research direction. Combining local descriptors, texture information, and gradient-based features may provide more comprehensive representations of image content. Integrating dimensionality reduction and feature selection with classifier-specific evaluation may further improve classification performance and robustness.

7. CONCLUSION

This paper reviewed feature extraction and feature selection techniques for image classification, focusing on SIFT, SURF, HOG, local descriptors, dimensionality reduction, and feature selection. The analysis demonstrates that each feature representation provides different advantages depending on the visual characteristics and requirements of the classification task. SIFT and SURF emphasize distinctive local structures, HOG captures shape and gradient information, while LBP and Gabor features provide complementary texture representations.

Dimensionality reduction and feature selection are essential for managing high-dimensional feature spaces and reducing redundant information. However, their effectiveness depends on preserving discriminative characteristics that contribute directly to classification performance. Ultimately, effective image classification requires an appropriate balance between feature representation quality, dimensionality, feature relevance, computational efficiency, and classifier performance.

REFERENCES

[1] 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.

[2] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, Jun. 2008, doi: 10.1016/j.cviu.2007.09.014.

[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] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, Jul. 2002, doi: 10.1109/TPAMI.2002.1017623.

[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] I. T. Jolliffe, "Choosing a Subset of Principal Components or Variables," in *Principal Component Analysis*, New York: Springer-Verlag, 2006, pp. 111–149. doi: 10.1007/0-387-22440-8_6.

[7] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, 2004, pp. 506–513. doi: 10.1109/cvpr.2004.1315206.
[8] A. Elisseeff and I. Guyon, "An Introduction to Variable and Feature Selection.," *Journal of Machine Learning Research*, vol. 3, no. Mar, pp. 1157–1182, 2003, [Online]. Available: http://www.jmlr.org/papers/v3/guyon03a.html
[9] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005, doi: 10.1109/TPAMI.2005.159.
[10] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005, doi: 10.1109/TPAMI.2005.188.
[11] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the seventh IEEE international conference on computer vision*, IEEE, 1999, pp. 1150–1157. doi: 10.1109/ICCV.1999.790410.
[12] J. Wu, Z. Cui, V. S. Sheng, P. Zhao, D. Su, and S. Gong, "A comparative study of SIFT and its variants," *Measurement Science Review*, vol. 13, no. 3, pp. 122–131, Jun. 2013, doi: 10.2478/msr-2013-0021.