Feature selection in image analysis: a survey

Verónica Bolón-Canedo · Beatriz Remeseiro

Abstract Image analysis is a prolific field of research which has been broadly studied in the last decades, successfully applied to a great number of disciplines. Since the apparition of Big Data, the number of digital images is explosively growing, and a large amount of multimedia data is publicly available. Not only is it necessary to deal with this increasing number of images, but also to know which features extract from them, and feature selection can help in this scenario. The goal of this paper is to survey the most recent feature selection methods developed and/or applied to image analysis, covering the most popular fields such as image classification, image segmentation, etc. Finally, an experimental evaluation on several popular datasets using well-known feature selection methods is presented, bearing in mind that the aim is not to provide the best feature selection method, but to facilitate comparative studies for the research community.

Keywords Feature selection · Image analysis · Pattern recognition · High dimensionality · Image datasets

1 Introduction

One of the most prolific fields of research is image analysis, which in fact has been broadly studied in the last few years. Part of its success is due to the fact that it can be applied to different disciplines, obtaining satisfactory results. A huge amount of information is contained in digital images, which is used with different purposes and extracted accordingly. For example, a manufacturing system of an aircraft factory needs to discard the pieces which present imperfections, so they could be

Verónica Bolón-Canedo
Department of Computer Science, Universidade da Coruña. Campus de Elviña s/n, 15071 A Coruña, Spain.
E-mail: vbolon@udc.es

Beatriz Remeseiro
Department of Computer Science, Universidad de Oviedo. Campus de Gijón s/n, 33203 Gijón, Spain.
E-mail: bremeseiro@uniovi.es

categorized into defective and non-defective pieces. Another example could be a collection of images of the Earth planet acquired from a satellite, and the necessity of partitioning them into several regions, such as water, forest, etc.

Different techniques have been developed to automatically analyze images, not only for general purposes [36,75] but also to be applied to very specific problems [10,78]. They are usually focused on the image processing side, i.e. on the extraction of relevant image properties which may include shape, color, texture, spatial information, etc. However, the availability of a large amount of images for most real-life problems, in conjunction with the large dimensionality of data extracted from certain types of images, make necessary the use of feature selection methods.

Feature selection (FS) is a popular preprocessing step which aims at identifying the relevant features and, at the same time, removing the redundant and irrelevant features. In this way, it is possible to scale down the dimensionality of the data without incurring in degradation in performance (and in some cases it is even possible to improve performance) [32]. These characteristics make FS an active research area, widely applied to real problems, mostly belonging to classification although it can be also applied to other fields such as regression, etc.

FS can help to deal with image analysis, and in fact it has been gaining importance in the last few years, not only to diminish the input dimensionality, but also to alleviate the computational burden required for extracting information from the images [86] or to understand the causes of disagreement among experts on the analysis of images for diagnosing diseases [3].

This work aims at offering the reader a broad review of the existing FS methods which have been applied (and in some cases developed ad-hoc) in the last few years to help in image analysis. Moreover, we have introduced an explanation of the different categories in this field and the datasets used. We have also designed an experimental study to try the effectiveness of FS on image datasets. We use four popular datasets, where some of them are binary and other have up to 26 classes, covering image classification and segmentation. We applied four state-of-the-art FS methods combined with five popular classifiers, leading to interesting conclusions and guidelines.

The rest of this manuscript is structured as follows. Section 2 is devoted to providing some necessary background about image analysis, specially for those machine learning experts who are not familiar with certain terms. Section 3 describes some basic concepts related to FS, and some popular techniques. Recent works of image analysis which benefit from FS are presented in Section 4. Some image datasets and repositories can be found in Section 5, and the experimental study is presented in Section 6. Finally, Section 7 concludes the manuscript.

2 Image analysis

Digital imaging has revolutionized the process of image acquisition, storage or transmission. The availability of digital images along with the Internet have noticeably increase the size of digital image collection [67]. As a result, many general-purpose imaging techniques have been developed over the last decades.

When talking about digital image processing techniques, it is typical to take into account three different techniques [28]: low-level, mid-level, and high-level. Low-level approaches work at the pixel/voxel level, using primitive operators, and

with images as input and output data. Mid-level methods include more elaborated tasks with images as input data, whilst the output data can be a set of characteristics/descriptors derived from images. High-level techniques concern the interpretation of the image content, trying to emulate the essential functionality of the human vision system. These three types of processes can be associated, respectively, with three different areas: image processing, image analysis, and computer vision —a brand of artificial intelligence. This paper is focused on the field of image analysis, also known as image understanding.

2.1 Terms

There are many different methods to automatically analyze images, each of them useful for a particular task or range of tasks. Therefore, image analysis techniques can be grouped into different fields, among which there are no clear-cut boundaries. Once reviewed the most up-date methods for FS applied to image analysis, four of these fields haven been considered in this review, and subsequently described.

2.1.1 Image classification

It consists in classifying an image, or a region of it, into a class from a set of classes according to the characteristics that it contains (see Figure 1). This categorization can be performed by using different approaches: template matching, which compares the image/region with a pattern; tree search algorithms; or more complex classifiers, which use a set of features. Note that digital image classification can be viewed as the fundamental prerequisite for other image analysis pursuits [77].
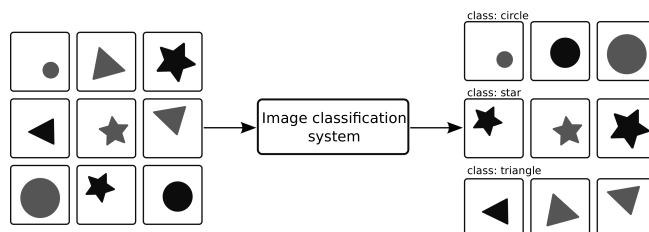
Fig. 1 Example of a system that classifies the images from a database into three classes.

2.1.2 Image segmentation

Its goal is to detect boundaries and objects in digital images, and can be defined as the identification of homogeneous regions (see Figure 2). That is, an image is divided in multiple regions so that each one is homogeneous according to some characteristic, whilst the union of any two or more segmented areas is not [13].
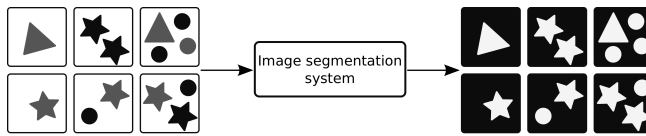
Fig. 2 Example of a system that segments different objects in the images from a database.

2.1.3 Image annotation

It consists in automatically assigning keywords or labels to a digital image (see Figure 3). The idea is to use some low-level features from the image, and then use a classifier to annotate it with concept labels [109]. Therefore, image annotation is also seen as a multi-class image classification task with a large number of class labels. Notice that automatic image annotation can be used as part of an image retrieval process with the aim of locating images in a database.
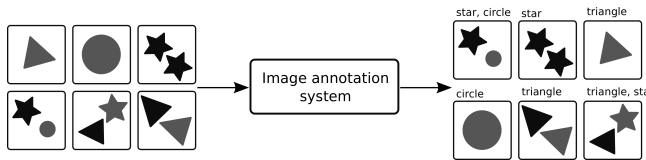

Fig. 3 Example of a system that labels the images from a database.

2.1.4 Image retrieval

Its objective is to search and retrieve images of interest from a database (see Figure 4). Two different frameworks can be found in the literature [67]: one is content-based, i.e. images are retrieved based on their visual content; and the other one is text-based, i.e. images are retrieved based on text descriptors manually annotated. In the second case, image retrieval is highly related to image annotation as previously stated.
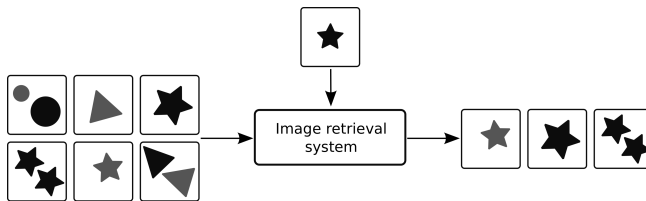

Fig. 4 Example of a system that retrieves the "star" images from a database.

2.2 First attempts

Image analysis techniques began to deal with different problems several decades ago, and a very large number of approaches can be found in the literature regarding the four fields previously presented. Image classification has been used in different problems using a great range of image features, such as spectral and spatial information [53], or texture [36]; and by means of several algorithms, such as neural networks [57], or decision trees [76]. In a similar way, different properties have been considered to perform image segmentation, including color [64], texture [8], or even a combination of them [93]. Regarding image annotation, texture information has been also commonly found in the literature [82], as well as ontology-based approaches [91]. Finally, image retrieval techniques are also based on properties obtained from image content, including color and shape [38].

2.3 Related works

There are other surveys about image analysis techniques although, to the best of our knowledge, none of them has been devoted to specifically analyzing the existing works on FS in this research field. Lu & Weng [70] presented a review of image classification techniques aimed at improving their performance, as well as discussing important issues that might have effect on it. Zhang et al. [109] provided a description of the latest developments in image retrieval as well as a complete review of automatic image annotation methods. In [18], Datta et al. surveyed a great amount of relevant contributions, both theoretical and empirical, regarding the automatic processes of image retrieval and annotation. Within the field of image segmentation, there are several works that review the most popular techniques, such as [110], which is focused on unsupervised methods; or [85], devoted to prediction.

3 Feature selection

As mentioned in the Introduction, the advent of Big Data problems, specially those containing a high number of features, has brought an important need for methods that can effectively reduce the dimensionality of the data. There are two main approaches to solve this problem: *feature extraction* and *feature selection* (FS). Basically, the main difference between them is that FS methods work by selecting a subset of the *original* features, while feature extraction methods are able to reduce the dimensionality by combining the existing features and creating *new* ones (see Figure 5). In the figure we can see an example in which we are trying to guess the race of a given person. A FS system would select those features that can help to determine the race, such as height, hair color, eye color, skin color or name. On the contrary, a feature extraction system would generate —for example— two new features, distinct from the original ones, that are likely to be combining information contained on the features selected by the FS system, but in this way we are losing the original meaning.

Another example, more related with image analysis, can be seen in Figure 6. In this case, in the left side we can see how the pixels (features) relevant for
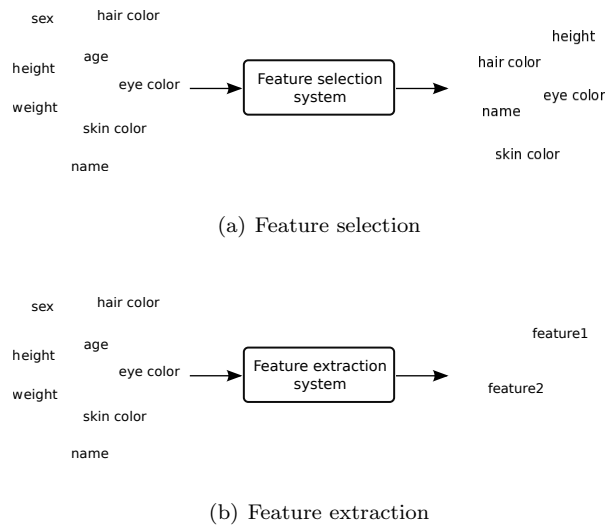
(a) Feature selection

(b) Feature extraction

Fig. 5 Example of a feature selection vs. feature extraction system

distinguishing between a four and nine digit are selected by a FS method and marked as blue dots and, in the right side, we can see the two principal components after a feature extraction process.
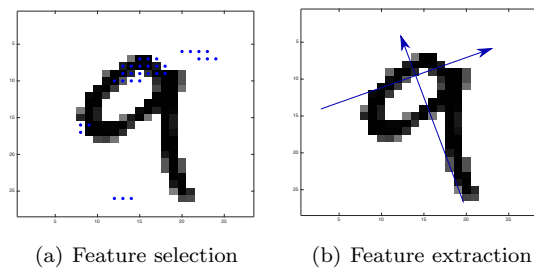

(a) Feature selection   (b) Feature extraction

Fig. 6 Feature selection vs. feature extraction on a image

Both techniques have their merits and demerits [114]. On the one hand, feature extraction methods have the power to generate a new set of features, which are usually more compact and of stronger discriminating power. This approach is the preferred one in applications in which model accuracy is very important, at the expense of interpretability. Examples of this type of applications are information retrieval, image analysis or signal processing. On the other hand, FS methods maintain the original features, so they are usually more adequate for data mining problems, such as text mining or genetic analysis, in which the original features are extremely important for model interpretability and knowledge extraction. More-

over, they offer the possibility of gaining speed and reducing costs (because in the future, the non-relevant features do not need to be considered).

Feature extraction is the preferred approach when dealing with image analysis [99, 87, 113, 107], and therefore there are plenty of works analyzing the application of these methods to this field [104, 30, 74, 44, 80]. However, although not so common, there are also image analysis works that employ FS methods as a preprocessing step, and this article is focused on reviewing them.

Existing FS methods are typically classified into three approaches: *filters*, *wrappers* and *embedded* methods [32, 4]. Filters work by using only the general characteristics of the datasets, with no influence of any classification model. For example, there is a huge amount of filters based on Information Theory that use mutual information as a measure of relevance for the features [7]. Many mutual information feature selection methods have been proposed in the last 25 years [101], and still new information theoretic feature selection methods are being proposed every year. Recent examples are a class-specific mutual information variation method for feature selection [25] and another method that integrates two groups of feature evaluation criteria [26], showing very competitive results. Different than filters, wrappers and embedded methods build the subset of relevant features based on a learning algorithm; the former by letting the prediction model to score the quality of the features for the prediction task and, the latter, by performing the selection of the features in the training process of a classifier. Given this relationship with the learner, wrapper and embedded methods are usually more accurate in their selection, but of course more costly in computational terms than filters. Moreover, wrappers are known for having the risk of over-fitting when the number of features is greater than that of samples [69], so they must be avoided in that situation. One of the most popular ensemble methods is SVM-RFE (Recursive Feature Elimination for Support Vector Machine) [33], which computes the importance of the features in the process of training a SVM. As for wrappers, they basically consist in using a learning algorithm (such as a SVM, decision tree, neural network, etc.) combined with a selection strategy such as forward or backward elimination.

Each year, new FS methods have been appearing belonging to the three categories mentioned above. However, this abundance of FS algorithms has not facilitated the choice of a particular method in a given situation, but quite the opposite. Nevertheless, despite the big amount of FS methods available, some of them have been able to stand out and their use has become very popular among researchers. Some of most popular ones are subsequently described.

- *Correlation-based feature selection* (CFS) [34] is a multivariate filter that chooses subsets of attributes uncorrelated among them but showing a high correlation with the class.
- *Consistency-based filter* [17] is a multivariate technique that also selects subsets of features but, in this case, according to the level of consistency with the class. It uses an inconsistency criterion to determine the acceptable data reduction rate.
- *Information Gain* [35] is a simple univariate filter that computes the mutual information between each attribute and the class, then providing an ordered ranking of all the attributes.
- *ReliefF* [48] is a popular multivariate filter extending a previous version (Relief [46]), which is based on nearest neighbors. The idea is to randomly pick an

example from the data and then find its nearest *hit* (neighbor from the same class) and *miss* (neighbor from the opposite class). After comparing the values of the selected instance with its hit and miss, the relevance score for each feature is updated. The values of a good feature should be similar in samples from the same class, whilst different in samples from different classes.

4 Recent works: 2008 – present

In the last decade there has been an important number of FS methods applied to image analysis; both standard and add-hoc to solve specific problems. In this section we review some of these works, classifying them according to the fields defined in Section 2: image classification, image segmentation, image annotation, and image retrieval.

4.1 Image classification

Image classification is perhaps one of the most prolific fields of image analysis. Moreover, and probably because of its similarity with regular classification, FS has been extensively applied to preprocess the data, trying to achieve the same level of success as in general data classification.

Among the vast amount of works that can be found in the literature, some of them present comparisons of several FS methods employed prior to the image classification step. Laliberte et al. [52] evaluated three FS methods for their ability to perform object-based vegetation classification, based on classification trees, distance and optimization. Their experiments conclude that the best option for this problem was to use classification tree analysis. Another example of comparison can be found in [83], in which several FS methods based on sequential search are studied, in this case to be applied to the problem of supervised color texture classification.

Not only have different FS methods been compared in the literature, but also different datasets to validate new algorithms. For instance, Barbu et al. [1] introduced a new FS method with annealing that combines the sequential algorithm and the regularization technique in order to be suitable for big data learning. In fact, it was successfully tested both on real and artificial data, and the experimental results were comparable to state-of-the-art methods while being computationally very efficient and scalable. An example of unsupervised FS can be found in [62], in which the authors propose a new unsupervised FS scheme based on redundancy analysis and non-negative spectral clustering. The experimental study carried out includes nine image datasets with different types of data (faces, objects, handwritten digits), and demonstrates the effectiveness of the proposal. Other example is the novel unsupervised sparse FS method presented in [16]. Experiments on two public UCI datasets and also on a new medical endoscopic image dataset showed that, when compared with state-of-the-art methods, their proposed method was able to choose the most discriminative features. At the same time, it also assigned meaningful weights to the useful attributes. In [111], the authors proposed to use sparsity-inducing regularization to define a supervised FS method that selects the minimum number of features. The experimentation was carried out on

different datasets, some of them related to image analysis, and the obtained results demonstrated the effectiveness of their approach. Ma et al. [72] presented a two-step wrapper comprising both feature ranking and feature subset methods. Several benchmark image data were employed in the testing phase, obtaining results in which the mean overall accuracy was significantly higher than previous approaches.

Other interesting approaches are focused on ensemble learning, based on the idea of combining the output of several methods (so-called experts) instead of relying on a single method, that might not be appropriate for all scenarios and situations. Wen et al. [105] recently proposed a fast ensemble for FS based on AdaBoost. In particular, it combined the values of the features with the class label. The experimental results outperformed state-of-the-art methods and demonstrated to being able to reduce times in the FS process. In [49], boosting meta-learning was used to select the most representative subset of features in a novel approach for visual object classification. The experiments demonstrated that accuracy and training time could be improved.

Incremental FS also appears in the specialized literature, as in [41], in which a new FS algorithm has demonstrated to beat previously published results on the CIFAR-10 dataset, employing for this a much smaller number of features than other methods. In the field of ant colony optimization there are some works devoted to image FS [11,9], which were able to improve classification results, using less features than other methods and reducing the processing times. A novel FS method was proposed in [117] for general image classification. It considers human factors and leverages the value of eye tracking data to find a subset of relevant input attributes, which are subsequently refined by a hybrid method combining FS based on mutual information (mRMR) and on SVMs. Experimentation carried out with two reference databases demonstrated that eye tracking data are of high relevance for FS when classifying images.

In the past few years, and specially with the arrival of Big Data, new challenges are to be faced by researchers to be able to deal with massive volumes of data. This huge amount of data makes that classification becomes a more complex and computationally demanding task, requiring the use of advanced techniques for Deep Learning [55] and, more specifically when using images, of Convolutional Neural Network (CNN) approaches [50,98]. Some works are recently appearing that apply FS based on deep learning to image classification. For example, Zou et al. [119] presented a new FS method based on deep learning formulated as a feature reconstruction problem. This new filter works by selecting those attributes that can be reconstructed as relevant and was successfully applied to remote sensing scene images. Roffo et al. [88] presented a new filter called infinite feature selection (Inf-FS) which performs the ranking step in an unsupervised manner. Their approach is based on affinity graphs and construct cost matrices that take into account pairwise relationships between attributes. They tested their approach on features extracted with a CNN from object recognition benchmark datasets, achieving results similar to the state-of-the-art.

Also in the context of high-dimensional data, Zheng et al. proposed a novel FS procedure to deal with incomplete datasets [115]. Their idea lies in defining a robust FS framework by taking into account the influence of outliers. Experimentation was carried out using different incomplete datasets, both real and artificial, one of which corresponds to Internet Advertisements and includes, among others,

geometrical features computed from advertisement images. The results demonstrated the adequacy of the proposed method compared to other FS methods.

Finally, it is worth noting that there are also works in the literature devoted to a specific application. In particular, we found an important number of articles promoting the use of FS for the classification of hyperspectral imagery, since it contains enormous amounts of data and the challenge is to obtain higher accuracy without incurring computational inefficiency. In this situation, dimensionality reduction techniques are often applied. Jia et al. [39] proposed a hybrid approach combining feature extraction (wavelet transform) and feature selection (affinity propagation), and concluded that their method outperformed other approaches that tackle feature extraction or feature selection individually. The same author [40] proposed later a FS framework, demonstrating the advantages of this proposal when compared to standard methods in a complex scenario with only a few instances per class being labeled. Also with very limited labeled samples, Zhang et al. [112] introduced a new FS algorithm in the field of particle swarm optimization (PSO), showing promising results. Based as well on PSO, but in this case integrated with a genetic algorithm, a new FS method was proposed in [27] for dealing with this type of data. In [96] the authors proposed an approach to select useful and non redundant Gabor features, since their original dimension was huge. This method was based on Markov blanket and symmetrical uncertainty. Also, a correlation coefficient for FS was proposed in [84], which used in combination with an optimized SVM allowed to improve state-of-the-art methods in terms of accuracy and efficiency. Finally, a common trend in this domain seems to be to employ FS schemes based on support vector machines [79, 51].

Regarding filter methods, a common approach is to apply those related with the *information theory* field. For instance, Kerroum et al. [45] proposed a new method with the goal of creating digital thematic maps for cartography exploitation (in the context of multispectral image classification). The method was based on Gaussian mixture models and computed the mutual information between the attributes and the class. The experimental results demonstrated that the selected textural features were the best option to improve the classification performance. Other works apply more sophisticated FS methods, for example through the use of SVM and different kernels. The authors in [100] proposed a method which works by automatically optimizing a linear combination of kernels, each of them based on a different subset of attributes. The experimental results, performed in contextual, multi- and hyperspectral, and multisource remote sensing data classification, demonstrated that the method is capable of ranking the features according to their relevance without losing computational efficiency.

Based on fuzzy-rough sets, we can find some works in the literature which used fuzzy-rough FS trying to classify Mars image [95, 94] showing promising results that could help in future Mars rover missions, both for ground-based or on-board image classification.

4.2 Image segmentation

There are several algorithms found in the literature to deal with image segmentation, not only as theoretical methods but also applied to different real-life problems. In both cases, FS has been applied in order to improve the algorithm performance.

Regarding generic methods, a novel framework was presented in [58] to combine two different types of segmentation, bottom-up and top-down, into an energy function. Authors used supervised learning and a feature induction method for conditional random fields. From a large set of candidate features, the supervised learning method chooses those relevant to define the energy function. For this task, they proposed their own FS algorithm, an iterative process that selects those features that maximize a conditional log likelihood. The learned algorithm achieved the performance of the state-of-art methods, being tested on three different datasets. In [63], we can find an approach based on genetic programming (GP) in the field of image segmentation. They proposed three novel FS methods, covering both single- and multi-objective GP. The experimentation demonstrated that the proposed multi-objective methods provide feature subsets that, in combination with classical classification systems, are able to improve previous results while being less computationally expensive. Cheng et al. [14] presented a hierarchical FS method used as part of an image segmentation system implemented on GPUs, which also included a fusion strategy with learning. The experimental results demonstrated its main advantage in speed, with no degradation in performance. Additionally, the system can be used in different image segmentation tasks.

Popular FS methods, such as random forests, can be also found in image segmentation problems. They were used for object class segmentation in [92]. This research demonstrated that the use of random forests allows to combine different image features in such a way that the pixel-wise segmentation performance is improved.

Different real problems have been also addressed in the area of image annotation. One representative example is fingerprint analysis, very used by e.g. law enforcements agencies, in which obtaining a correct segmentation is a critical issue. Sankaran et al. [90] focused on the automatic segmentation of fingerprints with the aim of differentiating types of patterns: ridge and non-ridge. In this context, they presented a FS technique based on the Relief algorithm to analyze how different category features affect the segmentation. Additionally, a random decision forest was used to categorize the local patches into two target classes: background and foreground. The validation was carried out by using three public databases and demonstrated the adequacy of the method to the problem at hand.

Another real problem in which FS was used to increase the performance of an image segmentation procedure is rock analysis. In [81], segmentation and classification were carried out to estimate the mineral composition in rock images. The minimum Redundancy Maximum Relevance (mRMR) algorithm was used to select 14 from 36 features extracted from rock images. In mining applications, the monitoring of rock composition has to work in real-time, which can be achieved thanks to FS methods that allow a reduction in both the dimensional space and the computational time.

Most of FS approaches determine features that are appropriate for a given dataset, being the main target of [37] to overcome this issue. The authors proposed a FS method independent of the dataset considered by predefining the effective feature types based on reasonable facts and selecting the appropriate candidate features for each feature type. In this manner, the features selected from a single image can be used in image segmentation applied to satellite images. The obtained results improved the ones provided by well-known FS methods, according to dif-

ferent evaluation measures. Also for satellite image segmentation, Chen et al. [12] introduced a new semi-supervised FS method. It works by generating different feature views (obtaining when attributes are distributed into several disjoint sets). The idea is to evaluate features and select them within each view. The experimentation on a dataset of very high resolution satellite images demonstrated that their new method performs better than the traditional algorithms.

4.3 Image annotation

Automatic image annotation consists in, given an input image, assigning it a semantic label, or more. It plays a key role to increase the effectiveness of other image processes such as retrieval or analysis. It can be defined as a special case of image classification and, therefore, FS is commonly used in this context too.

With the main aim of improving the performance of automatic systems for image annotation, researchers have to focus their attention on suitable frameworks for several purposes, including image content representation, feature extraction, classification algorithms, and feature selection. Regarding the latter issue, a FS approach was presented in [43] with the aim of improving image annotation. In this case, the contribution of each image feature was measured by means of mutual information, and its performance was measured by means of a non-linear factor in the evaluation function.

As most of the features used in these systems may be noisy and/or redundant, FS was used in [73] to represent the images in a more compact and precise way, which implies an improvement in terms of performance. More specifically, the authors proposed a novel method with two appealing properties: selecting the relevant attributes with a sparsity-based model, and determining the share sub-space of original features, useful in multi-label problems. Experimentation demonstrated that the proposed method for FS is robust as well as adequate for web images, which usually have multi-labels. In [97], a sparse FS framework was presented. The validation procedure demonstrated that this novel technique is both effective and efficient, as well as suitable for large-scale image annotation.

Genetic algorithms (GA) force a natural selection to find optimal values of some function, and thus they are susceptible to be used for FS. Lu et al. [71] used color, texture and shape properties to represent low-level features in their image annotation procedure. In order to optimize the weights of feature vectors, they defined the fitness function using a GA and k-nearest neighbor accuracy. In the same context, the authors in [59] proposed a method for image annotation using Adaboost and a FS method based on GA. Their idea lies in generating and optimizing a set of feature subsets at each iteration of the Adaboost method by means of a GA. In this case, two different approaches of GA-based FS were analyzed. In addition to genetic algorithms, another evolutionary computation paradigm has been used for FS applied to image annotation: particle swarm optimization (PSO). A novel scheme for image annotation was presented in [42], based on an improved quantum PSO method for visual FS. Additionally, the performance of this approach was improved by applying a boosting-based ensemble strategy. The experimentation demonstrated that the proposed scheme is adequate for the problem at hand.

Much attention has been attracted by multi-task FS in the last few years, since it often outperforms single-task FS. In [61], the authors addressed the problem of semi-supervised multi-task FS as applied to social image annotation. For this purpose, the authors proposed to use manifold regularization in order to manage the great imbalance between labeled and unlabeled instances. The process consists in estimating a FS matrix by integrating the obtained information into a learning framework, resulting in a novel method that outperforms state-of-art techniques. In the same line of research, Zeng et al. [108] presented a novel semi-supervised multi-label FS model and apply it to the task of multimedia annotation. The idea lies in combining both semi-supervised and multi-label feature learning into a single framework. They applied the proposed method to both web page and image annotation, using several types of real-world multimedia datasets, and demonstrated its effectivenesses.

4.4 Image retrieval

Image retrieval systems have become a focus of research in the field of image analysis and machine vision, and FS has been successfully applied to them as can be found in the literature and following summarized.

FS can be applied because of the very large number of image features and image classes. Regarding the first issue, experiments to compare different features for image retrieval were presented in [20]. The paper includes an analysis of a large set of different features and a comparison of them on different tasks, such as photo and building retrieval. In order to determine how different features can be used in combination, the proposed method analyzes the correlation between them and gives some recommendations to select an appropriate set of features, depending on the type of data.

A common complication that deserves attention in image retrieval is the existing difference between what features we humans see in a particular image, and the semantic features we use when it comes to describe it. For example, Lotfabadi et al. [68] presented a method to extract useful features from a feature vector. It combines a SVM classifier with a fuzzy rough set based on mutual information. The method has been compared with traditional systems that use other techniques for dimensionality reduction, and the experimentation demonstrated its adequacy in terms of accuracy and robustness by providing results highly relevant to the content of an image query.

On the other hand, cross-modal retrieval has lately attracted much attention because of the widespread use of multi-modal data. In this problem, relevant objects of one type of data are retrieved by means of another type of data used as the query. Therefore, researchers have to deal with two main issues: the measure of relevance and coupled FS. Both problems were tackled in [102] by means of a novel joint learning framework. The second problem, that is the key point in this survey, was approached by a learning procedure that uses the l_2-norm penalties on the projection matrices. This procedure allows to simultaneously select relevant and discriminative features, outperforming the state-of-the-art results.

There are also specific applications in the field of image retrieval. In [60], a novel approach of content-based image retrieval was presented as applied to remote sensing images. The authors proposed a novel scheme that, inspired on FS, selects

in an adaptive way the adequate vantage point tree indexing when considering different feature spaces. In this manner, the system is able to increase the response speed as well as the retrieval quality. The traffic vehicle search in large databases was addressed in [118], in which the authors define a local descriptor based on the gradient quantity as well as the spatial gradient distribution of the feature. Then, they propose an adaptive FS method by combining feature distinctive degree and the priori information. The experimentation carried out showed that their proposal outperformed the standard algorithms. Other application found in the literature is Batik image retrieval [23], being Batik a unique fabric from Indonesia. The authors analyzed the performance of FS and reduction techniques on the batik retrieval process. Particularly, they used sequential forward floating selection (SFFS) that consists of a forward step and a conditional backward step. The experimental results demonstrated that SFFS can improve the processing time allowing the method to be 1800 times faster.

Regarding many motion data based applications, it should be highlighted the important role played by human motion retrieval. Wang et al. [103] presented an adaptive multi-view FS technique to deal with this problem. In a first step, they used linear regression in a local way trying to learn Laplacian graphs based on multiple views, that were then combined together to take advantage of the complementary information between different attributes. Then, an objective function was formulated as a trace ratio optimization problem to remove from the original feature representation those feature components that are either irrelevant or redundant. Experimentation performed on two datasets publicly available show that the method is sound and achieved a state-of-art performance for motion data retrieval, as well as being adequate for other real-world applications.

4.5 Summary

We have reviewed the recent works promoting the use of FS to deal with image analysis, grouped according to the field they were applied to. Additionally, we have analyzed the nature of the FS methods usually applied. In Table 1 we can see a summary of all the methods reviewed.

Figure 7 illustrates the distribution of the FS methods present in the literature review, according to different aspects. Firstly, we have analyzed if the FS methods were new methods or classical techniques applied to deal with a given problem (see Figure 7(a)). Regarding this, we can say that most of the works proposed new methods, including generic approaches for a certain image analysis category or ad-hoc frameworks to solve very specific problems. Other works use existing techniques, although in both cases the experimental results show the adequacy of using FS in this domain.

Although many proposed methods are tested on benchmark and state-of-the-art datasets, others are applied to solve real problems. We have seen that analysis of remote sensing images through the use of feature selection methods has been the focus of much attention, especially when it comes to classification but also in image retrieval and image segmentation. Hyperspectral imagery, medical data or advertisements are other real applications that have been found in the field of image classification. Regarding image segmentation, feature selection has been used to deal with fingerprint analysis and rock analysis. For image annotation,

Table 1 Summary of recent works analyzed in Section 4.

Reference	Category	Development	Nature	Type
Laliberte et al. [52]	classification	classic	other	embedded
Porebski et al. [83]	classification	classic	other	filter
Jia et al. [39]	classification	other	other	filter
Jia et al. [40]	classification	new-generic	IT; hybrid	filter
Zhang et al. [112]	classification	new-generic	evo	filter
Ghamisi & Benediktsson [27]	classification	new-generic	evo; hybrid	embedded
Shen et al. [96]	classification	new-generic	IT	filter
Qi et al. [84]	classification	new-generic	IT	filter
Pal & Foody [79]	classification	classic	IT; ens; svm; other	filter; embedded
Kuo et al. [51]	classification	new-generic	svm	wrapper
Kerroume et al. [45]	classification	new-generic	IT	filter
Tuia et al. [100]	classification	new-generic	svm	embedded
Barbu et al. [1]	classification	new-generic	evo	embedded
Li & Tang [62]	classification	new-generic	other	embedded
Cong et al. [16]	classification	new-generic	other	embedded
Zhang et al. [111]	classification	new-generic	other	embedded
Wen et al. [105]	classification	new-ad-hoc	ens	wrapper
Korytkowski et al. [49]	classification	new-generic	ens	embedded
Jia et al. [41]	classification	new-generic	other	wrapper
Shang et al. [95]	classification	classic	other	filter
Shang & Barnes [94]	classification	new-generic	other	filter
Chen et al. [11]	classification	new-generic	evo	wrapper
Chen et al. [9]	classification	new-generic	evo	wrapper
Zhou et al. [117]	classification	new-ad-hoc	hybrid	filter; wrapper
Ma et al. [72]	classification	new-generic	IT	wrapper
Zou et al. [119]	classification	new-ad-hoc	other	filter
Roffo et al. [88]	classification	new-generic	other	filter
Zheng et al. [115]	classification	new-generic	other	filter
Levin & Weiss [58]	segmentation	new-generic	other	filter
Liang et al. [63]	segmentation	new-generic	evo	filter
Schroff et al. [92]	segmentation	new-generic	ens	embedded
Cheng et al. [14]	segmentation	new-generic	svm	wrapper
Sankaran et al. [90]	segmentation	new-ad-hoc	other	filter
Perez et al. [81]	segmentation	classic	IT	filter
Izadipour et al. [37]	segmentation	new-ad-hoc	other	filter
Chen et al. [12]	segmentation	new-ad-hoc	other	filter
Jin et al. [43]	annotation	new-generic	IT	filter
Ma et al. [73]	annotation	new-generic	other	filter
Shi et al. [97]	annotation	new-generic	other	filter
Lu et al. [71]	annotation	new-generic	evo	wrapper
Li et al. [59]	annotation	new-generic	evo	wrapper
Jin & Jin [42]	annotation	new-generic	evo	filter
Li et al. [61]	annotation	new-generic	other	embedded
Zeng et al. [108]	annotation	new-generic	other	embedded
Deselaers et al. [20]	retrieval	new-generic	other	other
Lotfabadi et al. [68]	retrieval	new-generic	IT	filter
Wang et al. [102]	retrieval	new-generic	other	filter
Li et al. [60]	retrieval	new-ad-hoc	ens	embedded
Zhu et al. [118]	retrieval	new-ad-hoc	other	embedded
Fahmi et al. [23]	retrieval	classic	other	wrapper
Wang et al. [103]	retrieval	new-ad-hoc	other	filter

the real applications we found are mainly based on multimedia datasets. Finally, feature selection methods have been successfully applied in image retrieval to real problems such as traffic vehicle search or human motion retrieval.

We have also analyzed the nature of the FS methods (see Figure 7(b)). In this sense, we could note that the methods based on information theory are still in use, in spite of the fact that they have appeared several years ago —a good review about this kind of methods is provided by Vergara & Estevez [101]. En-

semble learning is an approach that combines the results of multiple methods (or experts) aiming at obtaining better performance than that of any single method. Mostly, this approach has been applied to classification problems, but also to image analysis, being Adaboost and Random Forest among the most used techniques. The evolutionary computing paradigm [106], comprising genetic algorithms and particle swarm optimization, is widely employed. Possibly because of the good outcomes of the SVM classifier, some FS methods based on it and its different kernels have been also used. It is also worth to mention the tendency to combine different methods, also known as hybrid methods, trying to improve the performance achieved with classical FS methods. Finally, other approaches found in the literature are sparse FS, wrappers, fuzzy-based approaches, regularization techniques, or methods based on distances.

Finally, we have analyzed the different types of FS methods according to the three approaches detailed in Section 3. As can be observed in Figure 7(c), most of the approaches correspond to filters, regardless of the category considered, and probably due to their independence of the classification model and their lower risk of overfitting. Embedded methods are also quite used in the literature, especially in the field of classification, probably because of their ability of performing feature selection and classification at the same time. Particularly popular are those embedded methods based on the ℓ_1-norm. Wrappers, possibly because of their high computational cost, are the less used approach.

Notice that, normally, deep learning algorithms are used to *extract* relevant features. By removing the last layer, one can take the final layer as a feature vector. These are the so-called *deep features* [116, 47, 66]. This is a feature extraction procedure so that is the reason why it is not reviewed in this article.
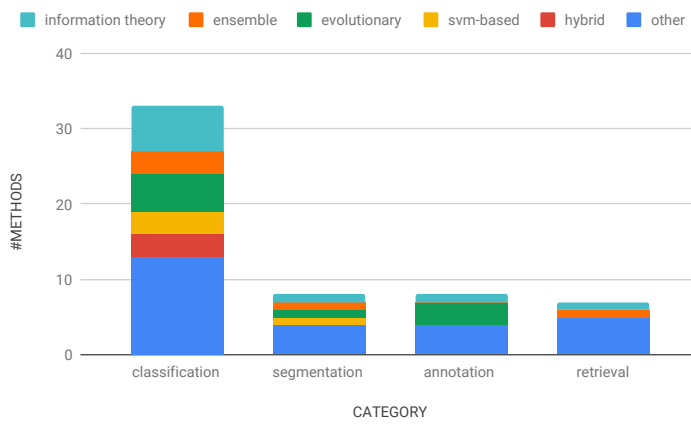
5 Image datasets

When researchers started to work in the field of image analysis, a key point was to find datasets publicly available to test their new approaches. Nowadays, several image databases are commonly used as benchmark datasets in different topics. Table 2 presents some of the most popular ones, and includes a brief description of them as well as some useful information.

Among all these image databases, it should be highlighted ImageNet [19] as one of the most popular collections. It is arranged based on the WordNet hierarchy, and each of its nodes is characterized by thousands of images. Currently, it has an average of over five hundred images per node, and more than 14 millions in total. Note that it became more popular thanks to the ImageNet Large Scale Visual Recognition Challenge (ILSVRC)[1], which helps image analysis researchers to test their new algorithms at a large scale. COCO is also well-known due to its challenges that include their joint organization with ILSVRC in 2016 and Coco + Places 2017. There are other databases that are also very popular due to similar challenges, such as for example PASCAL VOC and the PASCAL Visual Object Classes Challenge [21]. Although the VOC challenges have now finished (2005-2012), the PASCAL VOC Evaluation Server as well as the different versions of the PASCAL VOC dataset continue to be available for benchmarking.

---

[1] http://www.image-net.org/challenges/LSVRC/
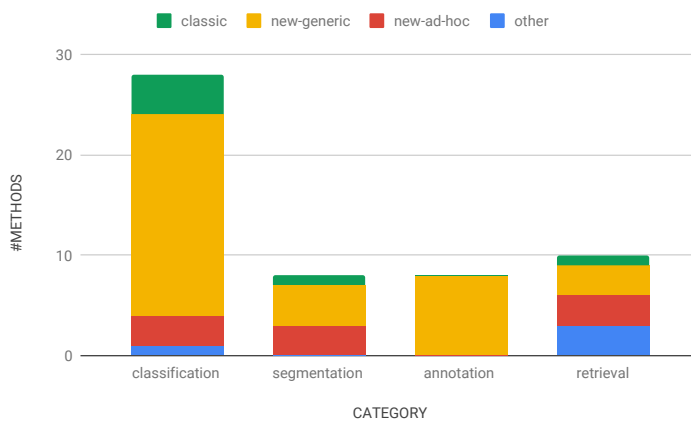
(a) Development

(b) Nature

(c) Type

Fig. 7 Distribution of feature selection approaches on the review of the literature (Section 4), according to their novelty, nature and type.

Table 2 Representative image datasets used as benchmarks in different fields.

ImageNet [19], Princeton and Standford Universities. [On-line] `http://www.image-net.org/` Description: image database organized following the Word-Net hierarchy, useful for several visual recognition tasks. Data provided: original images, image URLs, SIFT features, object bounding boxes, and object attributes.	# images: > 14M # classes: > 21K
COCO [65], Cornell Tech, MSR, Caltech, and others. [On-line] `http://mscoco.org/` Description: large-scale image database for object detection, segmentation and captioning. Data provided: original images, annotations, and some dataset tools.	# images: 330K # classes: 80
PASCAL VOC [22], University of Leeds and others. [On-line] `http://host.robots.ox.ac.uk/pascal/VOC/` Description: standardized image database for object class recognition. Data provided: original images, annotated objects, segmentations, and a set of tools for accessing them.	# images: > 11K # classes: 20
MNIST [56], New York Univ., Google Labs and Microsoft. [On-line] `http://yann.lecun.com/exdb/mnist/` Description: image database of handwritten digits corresponding to different writers, useful for image classification. Data provided: size-normalized and centered images (no preprocessing or formatting required).	# images: 70K # classes: 10
Caltech256 [29], California Institute of Technology. [On-line] `http://www.vision.caltech.edu/Image_Datasets/Caltech256/` Description: image database useful for object classification. Data provided: original images organized in a taxonomy, object annotations as outlines, and a Matlab script to visualize them.	# images: > 30K # classes: 256
LabelMe [89], MIT. [On-line] `http://labelme.csail.mit.edu/Release3.0/browserTools/php/dataset.php` Description: image database of outdoor pictures, useful for object recognition and image segmentation. Data provided: original images, and partially annotated objects.	# images: > 4K # classes: 12
NUS-WIDE [15], National University of Singapore. [On-line] `http://lms.comp.nus.edu.sg/research/NUS-WIDE.htm` Description: web image database useful for image annotation and retrieval. Data provided: image URLs, original images, tags, and six different types of low-level features.	# images: > 269K # classes: > 5K
Food-101 [6], ETH Zurich. [On-line] `https://www.vision.ee.ethz.ch/datasets_extra/food-101/` Description: image database useful for food classification, including different food categories and some noise in both images and labels. Data provided: original images, food categories.	# images: 101K # classes: 101
Labelled Faces in the Wild [54], Univ. of Massachusetts. [On-line] `http://vis-www.cs.umass.edu/lfw/` Description: image database of face photographs useful for face recognition. Data provided: original images, images aligned with deep funneling, and annotations with the person pictured.	# images: > 13K # classes: > 5K

Note that image datasets include, in general, a set of row images (samples) and their respective annotations (labels); in contrast to the datasets commonly used in other machine learning problems in which samples are directly represented by feature vectors. Therefore, when dealing with row images, some feature extraction is required as a previous step to apply FS. If researchers want to focus their efforts on the FS procedure, it should be more appropriate for them to directly manage a set of features or image properties. In this sense, some popular image datasets (e.g. ImageNet) already include a set of features computed from the images (e.g. SIFT features). Additionally, we should highlight the UCI Machine Learning Repository [2], which contains datasets for general machine learning purposes, some of which are related to image analysis. The datasets included in this repository are provided in a way that machine learning algorithms (such as FS methods) can be directly applied. For this reason, some of the image datasets included in the UCI repository will be used in Section 6 for experimentation.

6 An experimental study

Typically, the most important benefits from performing FS on image analysis are to improve the learning performance or to better understand what features or pixels are important. If we are interested in class prediction, it is necessary to employ afterwards a supervised machine learning technique, such as a classifier. On the contrary, if the goal is data understanding, the classification part is ignored and the selected features have to be individually evaluated. In this section we present experiments focused on class prediction.

6.1 Datasets

We have chosen four image datasets from the UCI repository (see Section 5). The reason to choose these datasets is because they have the features already extracted, so we do not introduce another layer of complexity depending on the method used to extract the features from the image. In Table 3 we can see a summary of the main characteristics of these datasets.

Table 3 Summary description of the datasets used in the experimental study

Dataset	# Feats.	# Samp. Train	# Samp. Test	# Classes
Gisette	5000	6000	1000	2
Image segmentation	19	1540	770	7
Letter recognition	16	13334	6666	26
Semeion	256	1062	531	10

– Gisette is a handwritten digit recognition problem. The problem is to separate the highly confusible digits '4' and '9'. This dataset is one of five datasets of the NIPS 2003 feature selection challenge [31]. The digits were size-normalized and centered in a fixed-size image of dimension 28x28. The original data were modified for the purpose of the FS challenge. In particular, pixels were sampled

at random in the middle top part of the feature containing the information necessary to disambiguate 4 from 9 and higher order features were created as products of these pixels to plunge the problem in a higher dimensional feature space. They also added a number of distracting features called 'probes' having no predictive power. The order of the features and patterns was randomized.

– Image Segmentation is an image dataset described by high-level numeric-valued attributes. The instances were drawn randomly from a database of 7 outdoor images. The images were hand-segmented to create a classification for every pixel. Each instance is a 3x3 region.

– Letter Recognition is a database of character image features, in which the objective is to identify each of a large number of black-and-white rectangular pixel displays as one of the 26 capital letters in the English alphabet. The character images were based on 20 different fonts and each letter within these 20 fonts was randomly distorted to produce a file of 20000 unique stimuli. Each stimulus was converted into 16 primitive numerical attributes (statistical moments and edge counts), which were then scaled to fit into a range of integer values from 0 through 15.

– Semeion Handwritten Digit dataset was constructed by collecting 1593 hand-written digits from around 80 people, and then scan and stretch them in a 16×16 rectangular box with a gray scale of 256 values. Then each pixel of each image was scaled between 0 and 1, by setting to 0 every pixel whose value was under the value 127 of the gray scale (127 included) and setting to 1 each pixel whose original value in the gray scale was over 127. Finally, each binary image was scaled again into a 16×16 square box (the final 256 binary attributes). Each person wrote on a paper all the digits from 0 to 9, twice. The commitment was to write the digit the first time in the normal way (trying to write each digit accurately) and the second time in a fast way (with no accuracy).

6.2 Results

The goal of this section is to perform an experimental study using four representative image datasets extracted from the UCI Machine Learning Repository and some classical widely used FS methods, providing the readers with some baselines for their comparisons. For this purpose, we have chosen four popular FS algorithms that can be considered as state-of-the-art and extensively used by researchers in Machine Learning [4]: CFS, consistency-based, Information Gain and ReliefF. Two of the FS methods (CFS and consistency-based) return a subset of features, whilst the other two (Information Gain and ReliefF) provide an ordered ranking of the features. These methods were selected because they are available in popular tools used by researchers, such as Weka, Matlab, RapidMiner or KEEL. Among them, we have chosen Weka[2] since it includes the four of them and it is very easy to use even for non experienced researchers. Note that for the ranker methods, we show the performance when the top 40% of the features are retained.

In order to evaluate the adequacy of these methods over image data, five well-known classifiers were chosen: C4.5, naive Bayes (NB), Support Vector Machine (SVM), Random Forest (RF) and k-NN (with $k = 3$, in this case). Except for the

---

[2] `https://www.cs.waikato.ac.nz/ml/weka/downloading.html`

parameter k in the k-NN classifier, the rest of the parameters have the values given by default in Weka, for an easy reproducibility of the experiments. In the case of Gisette dataset, we used the original division into train and test sets, and for the remaining, we performed a holdout validation using 2/3 of data for training and 1/3 for test.

Table 4 shows the experimental results obtained by the different classifiers in the four datasets. As can be seen, the datasets Gisette, letter recognition and image segmentation are in general benefited from the application of FS methods, improving their classification accuracy with respect to using all the features. On the contrary, for Semeion the best results were obtained when using the whole set of features, which suggests that all the pixels are necessary to correctly determine the class. Analyzing in more detail the behavior of the FS methods as well as the influence of the classifier on the studied datasets, some interesting conclusions can be drawn:

- The best performances are obtained by Random Forest, kNN and, specially, SVM. This is not surprising since Random Forest and SVM are reported to be powerful classifiers [24]. Notice the bad results obtained by SVM on the letter recognition dataset compared with the performance of kNN. This can be due to the fact that this dataset has a high number of classes (26) and SVMs are designed to deal with binary problems, so they have to perform one-versus-rest or one-versus-one approaches, leading to loss of accuracy.
- Focusing on the FS methods, in general for all datasets, the subset filters (CFS and consistency-based) show an outstanding behavior. The poor performance of the ranker methods (except for Gisette dataset) can be explained by the restriction of having to establish a threshold for the number of features to keep, which might not be enough in some cases. In the case of subset filters, the number of selected features is supposed to be the optimal one for a given dataset. Thus, the main disadvantage of rankers is the need for setting the threshold a priori, with the risk of choosing a too large or too small number.
- All the methods tested except Information Gain are multivariate, which in theory provide a better performance than univariate methods. In average for all the datasets and classifiers tested, in fact Information Gain obtains the worst test accuracy. However, there are some cases in which for Gisette dataset this method leads to the best classification accuracy, suggesting that there are not important interactions between features in this dataset.

In light of the above, it can be seen that the results obtained by this experimental study are highly dependent on the classifier, the FS method, and in particular the dataset. Although a detailed analysis of the results is outside the scope of this paper, the authors recommend the use of subset FS methods (in particular, CFS) in combination with SVM or Random Forest classifiers.

7 Conclusions

In this work we have provided an exhaustive review and analysis of the recent contributions of FS as a preprocessing step applied to the field of image analysis. Nowadays, with the Big Data phenomenon surrounding us, the necessity of using FS methods is more important than ever, although it was decades ago when image

Table 4 Classification accuracy (%) obtained by five classifiers in the four datasets considered. The best result of each classifier is marked in bold face. The best absolute result for each dataset is highlighted in shadowed.
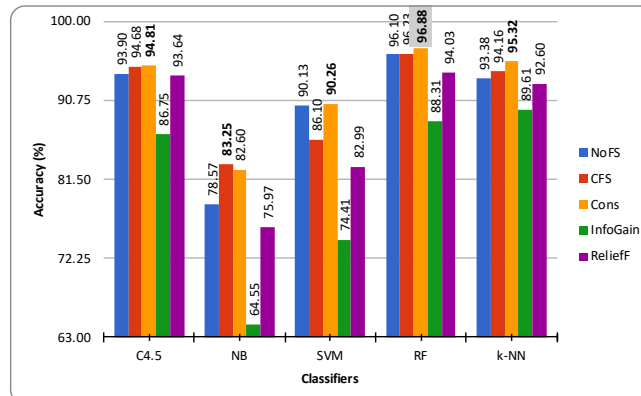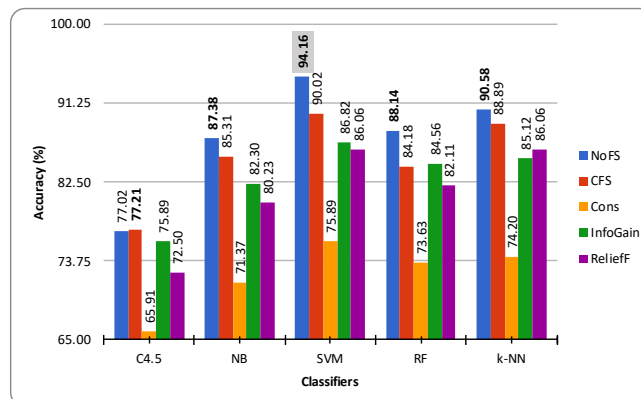

(a) Gisette
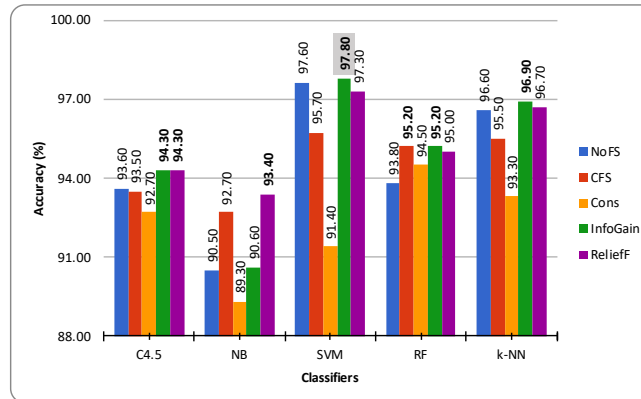
(b) Letter recognition
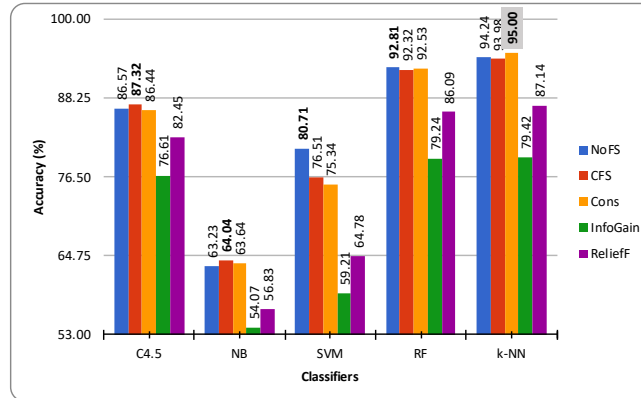
(c) Image recognition

(d) Semeion

analysis researchers noticed the need of knowing which features had to be extracted from each pixel [5].

Image analysis covers a wide field of applications and, thus, of specific techniques. This work has been focused on those in which FS has been mostly applied —image classification, image segmentation, image annotation, and image retrieval—, providing an extensive description of each of them to avoid confusions to the interested reader who is not an expert on the field. Analogously, basic FS concepts are also explained.

The goal of this work is to explain and review the different image analysis categories and the FS approaches that have been applied to them, bringing together as much up-to-date knowledge as possible. Thus, recent works found in the specialized literature have been exhaustively examined, in an attempt to describe the applications of FS to the different subfields of image analysis. Furthermore, the most popular data repositories in this field have been briefly presented.

Finally, we have performed a practical evaluation for FS methods using image datasets in which we analyze the results obtained. We chose four widely-used datasets to apply over them four classical FS methods. In order to obtain the final classification accuracy, five well-known classifiers were used. This set of experiments also aims at facilitating future comparative studies when a researcher proposes a new method.

Regarding the opportunities for future research, it is essential not to overlook the new scenario of Big Data, in which it is not only important to deal with millions of pixels in a given image, but also with millions of images at the time. As pointed out in [5], "data is being collected at an unprecedented fast pace and, consequently, needs to be processed rapidly". We live in a society where social media networks are everywhere, specially thanks to portable devices, which generate huge amounts of data each second. Therefore, we need sophisticated methods able to process millions of images in real time. To this end, on-line FS methods are in need, which still remain a challenge for researchers. Moreover, another way to solve this issue is to develop distributed FS methods, trying to alleviate the computational burden required for processing large amounts of images.

References

1. Barbu, A., She, Y., Ding, L., Gramajo, G.: Feature Selection with Annealing for Computer Vision and Big Data Learning. IEEE Transactions on Pattern Analysis and Machine Intelligence **39**(2), 272–286 (2017)
2. Blake, C.L., Merz, C.J.: UCI Machine Learning Repository. Department of Information and Computer Science, University of California [Online]. Available: http://archive.ics.uci.edu/ml/ **55** (1998)

3. Bolón-Canedo, V., Ataer-Cansizoglu, E., Erdogmus, D., Kalpathy-Cramer, J., Fontenla-Romero, O., Alonso-Betanzos, A., Chiang, M.: Dealing with inter-expert variability in retinopathy of prematurity: A machine learning approach. Computer Methods and Programs in Biomedicine **122**(1), 1–15 (2015)
4. Bolon-Canedo, V., Sanchez-Marono, N., Alonso-Betanzos, A.: Feature selection for high-dimensional data. Springer (2015)
5. Bolón-Canedo, V., Sánchez-Maroño, N., Alonso-Betanzos, A.: Recent advances and emerging challenges of feature selection in the context of big data. Knowledge-Based Systems **86**, 33–45 (2015)
6. Bossard, L., Guillaumin, M., Van Gool, L.: Food-101 – Mining Discriminative Components with Random Forests. In: European Conference on Computer Vision, pp. 446–461 (2014)
7. Brown, G., Pocock, A., Zhao, M.J., Luján, M.: Conditional likelihood maximisation: a unifying framework for information theoretic feature selection. Journal of machine learning research **13**(Jan), 27–66 (2012)
8. du Buf, J.M.H., Kardan, M., Spann, M.: Texture feature performance for image segmentation. Pattern Recognition **23**(3–4), 291–309 (1990)
9. Chen, B., Chen, L., Chen, Y.: Efficient ant colony optimization for image feature selection. Signal processing **93**(6), 1566–1576 (2013)
10. Chen, E.L., Chung, P.C., Chen, C.L., Tsai, H.M., Chang, C.I.: An automatic diagnostic system for CT liver image classification. IEEE Transactions on Biomedical Engineering **45**(6), 783–794 (1998)
11. Chen, L., Chen, B., Chen, Y.: Image feature selection based on ant colony optimization. In: AI 2011: Advances in Artificial Intelligence, pp. 580–589 (2011)
12. Chen, X., Liu, W., Su, F., Shao, G.: Semi-supervised multiview feature selection with label learning for VHR remote sensing images. In: IEEE International Geoscience and Remote Sensing Symposium, pp. 2372–2375 (2016)
13. Cheng, H.D., Jiang, X., Sun, Y., Wang, J.: Color image segmentation: advances and prospects. Pattern Recognition **34**(12), 2259–2281 (2001)
14. Cheng, M.M., Liu, Y., Hou, Q., Bian, J., Torr, P., Hu, S.M., Tu, Z.: HFS: Hierarchical Feature Selection for Efficient Image Segmentation. In: European Conference on Computer Vision, pp. 867–882 (2016)
15. Chua, T.S., Tang, J., Hong, R., Li, H., Luo, Z., Zheng, Y.: NUS-WIDE: a real-world web image database from National University of Singapore. In: ACM International Conference on Image and Video Retrieval, p. 48 (2009)
16. Cong, Y., Wang, S., Fan, B., Yang, Y., Yu, H.: UDSFS: unsupervised deep sparse feature selection. Neurocomputing **196**, 150–158 (2016)
17. Dash, M., Liu, H.: Consistency-based search in feature selection. Artificial intelligence **151**(1), 155–176 (2003)
18. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Image retrieval: Ideas, influences, and trends of the new age. ACM Computing Surveys **40**(2), 5 (2008)
19. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255 (2009)
20. Deselaers, T., Keysers, D., Ney, H.: Features for image retrieval: an experimental comparison. Information Retrieval **11**(2), 77–107 (2008)
21. Everingham, M., Eslami, S.A., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The Pascal Visual Object Classes Challenge: A Retrospective. International Journal of Computer Vision **111**(1), 98–136 (2015)
22. Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The PASCAL Visual Object Classes (VOC) Challenge. International Journal of Computer Vision **88**(2), 303–338 (2010)
23. Fahmi, H., Zen, R.A., Sanabila, H.R., Nurhaida, I., Arymurthy, A.M.: Feature Selection and Reduction for Batik Image Retrieval. In: Proceedings of the Fifth International Conference on Network, Communication and Computing, pp. 47–52 (2016)
24. Fernández-Delgado, M., Cernadas, E., Barro, S., Amorim, D.: Do we need hundreds of classifiers to solve real world classification problems. Journal of Machine Learning Research **15**(1), 3133–3181 (2014)
25. Gao, W., Hu, L., Zhang, P.: Class-specific mutual information variation for feature selection. Pattern Recognition **79**, 328–339 (2018)

26. Gao, W., Hu, L., Zhang, P., Wang, F.: Feature selection by integrating two groups of feature evaluation criteria. Expert Systems with Applications **110**, 11–19 (2018)
27. Ghamisi, P., Benediktsson, J.A.: Feature selection based on hybridization of genetic algorithm and particle swarm optimization. IEEE Geoscience and Remote Sensing Letters **12**(2), 309–313 (2015)
28. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, Third edition. Pearson Prentice Hall (2008)
29. Griffin, G., Holub, A., Perona, P.: Caltech-256 object category dataset (2007)
30. Guo, G., Fu, Y., Dyer, C.R., Huang, T.S.: Image-based human age estimation by manifold learning and locally adjusted robust regression. IEEE Transactions on Image Processing **17**(7), 1178–1188 (2008)
31. Guyon, I., Gunn, S., Ben-Hur, A., Dror, G.: Result analysis of the NIPS 2003 feature selection challenge. In: Advances in Neural Information Processing Systems, pp. 545–552 (2005)
32. Guyon, I., Gunn, S., Nikravesh, M., Zadeh, L.A.: Feature Extraction: Foundations and Applications. Springer Verlag (2006)
33. Guyon, I., Weston, J., Barnhill, S., Vapnik, V.: Gene selection for cancer classification using support vector machines. Machine learning **46**(1-3), 389–422 (2002)
34. Hall, M.A.: Correlation-based feature selection for machine learning. Ph.D. thesis, The University of Waikato (1999)
35. Hall, M.A., Smith, L.A.: Practical feature subset selection for machine learning. Computer Science **98**, 181–191 (1998)
36. Haralick, R.M., Shanmugam, K., Dinstein, I.: Texture Features for Image Classification. IEEE Transactions on Systems, Man, and Cybernetics **3**, 610–621 (1973)
37. Izadipour, A., Akbari, B., Mojaradi, B.: A Feature Selection Approach for Segmentation of Very High-Resolution Satellite Images. Photogrammetric Engineering & Remote Sensing **82**(3), 213–222 (2016)
38. Jain, A.K., Vailaya, A.: Image retrieval using color and shape. Pattern Recognition **29**(8), 1233–1244 (1996)
39. Jia, S., Qian, Y., Li, J., Liu, W., Ji, Z.: Feature extraction and selection hybrid algorithm for hyperspectral imagery classification. In: IEEE International Geoscience and Remote Sensing Symposium, pp. 72–75 (2010)
40. Jia, S., Zhu, Z., Shen, L., Li, Q.: A two-stage feature selection framework for hyperspectral image classification using few labeled samples. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **7**(4), 1023–1035 (2014)
41. Jia, Y., Huang, C., Darrell, T.: Beyond spatial pyramids: Receptive field learning for pooled image features. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3370–3377 (2012)
42. Jin, C., Jin, S.W.: Automatic image annotation using feature selection based on improving quantum particle swarm optimization. Signal Processing **109**, 172–181 (2015)
43. Jin, C., Liu, J., Guo, J.: A Hybrid Model Based on Mutual Information and Support Vector Machine for Automatic Image Annotation. In: Artificial Intelligence Perspectives and Applications, pp. 29–38 (2015)
44. Juan, L., Gwun, O.: A comparison of SIFT, PCA-SIFT and SURF. International Journal of Image Processing **3**(4), 143–152 (2009)
45. Kerroum, M.A., Hammouch, A., Aboutajdine, D.: Textural feature selection by joint mutual information based on gaussian mixture model for multispectral image classification. Pattern Recognition Letters **31**(10), 1168–1174 (2010)
46. Kira, K., Rendell, L.A.: The feature selection problem: Traditional methods and a new algorithm. In: National Conference on Artificial Intelligence, pp. 129–129 (1992)
47. Kong, T., Yao, A., Chen, Y., Sun, F.: Hypernet: Towards accurate region proposal generation and joint object detection. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 845–853 (2016)
48. Kononenko, I.: Estimating attributes: analysis and extensions of RELIEF. In: Machine Learning: ECML-94, pp. 171–182 (1994)
49. Korytkowski, M., Rutkowski, L., Scherer, R.: Fast image classification by boosting fuzzy classifiers. Information Sciences **327**, 175–182 (2016)
50. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp. 1097–1105 (2012)

51. Kuo, B.C., Ho, H.H., Li, C.H., Hung, C.C., Taur, J.S.: A kernel-based feature selection method for SVM with RBF kernel for hyperspectral image classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **7**(1), 317–326 (2014)
52. Laliberte, A.S., Browning, D., Rango, A.: A comparison of three feature selection methods for object-based classification of sub-decimeter resolution UltraCam-L imagery. International Journal of Applied Earth Observation and Geoinformation **15**, 70–78 (2012)
53. Landgrebe, D.A.: The development of a spectral-spatial classifier for earth observational data. Pattern Recognition **12**(3), 165–175 (1980)
54. Learned-Miller, E., Huang, G.B., RoyChowdhury, A., Li, H., Hua, G.: Labeled faces in the wild: A survey. In: Advances in Face Detection and Facial Image Analysis, pp. 189–248 (2016)
55. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015)
56. LeCun, Y., Cortes, C., Burges, C.J.: MNIST handwritten digit database. AT&T Labs [Online]. Available: http://yann. lecun.com/exdb/mnist **2** (2010)
57. Lee, J., Weger, R., Sengupta, S., Welch, R.: A neural network approach to cloud classification. IEEE Transactions on Geoscience and Remote Sensing **28**(5), 846–855 (1990)
58. Levin, A., Weiss, Y.: Learning to combine bottom-up and top-down segmentation. International Journal of Computer Vision **81**(1), 105–118 (2009)
59. Li, R., Lu, J., Zhang, Y., Zhao, T.: Dynamic Adaboost learning with feature selection based on parallel genetic algorithm for image annotation. Knowledge-Based Systems **23**(3), 195–201 (2010)
60. Li, S., Yu, H., Yuan, L.: A Novel Approach to Remote Sensing Image Retrieval with Multi-feature VP-Tree Indexing and Online Feature Selection. In: IEEE Second International Conference on Multimedia Big Data, pp. 133–136 (2016)
61. Li, Y., Shi, X., Du, C., Liu, Y., Wen, Y.: Manifold regularized multi-view feature selection for social image annotation. Neurocomputing **204**, 135–141 (2016)
62. Li, Z., Tang, J.: Unsupervised feature selection via nonnegative spectral analysis and redundancy control. IEEE Transactions on Image Processing **24**(12), 5343–5355 (2015)
63. Liang, Y., Zhang, M., Browne, W.N.: Image feature selection using genetic programming for figure-ground segmentation. Engineering Applications of Artificial Intelligence **62**, 96–108 (2017)
64. Lim, Y.W., Lee, S.U.: On the color image segmentation algorithm based on the thresholding and the fuzzy c-means techniques. Pattern Recognition **23**(9), 935–952 (1990)
65. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft COCO: Common Objects in Context. In: European Conference on Computer Vision, pp. 740–755 (2014)
66. Liu, Y., Cheng, M.M., Hu, X., Wang, K., Bai, X.: Richer convolutional features for edge detection. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 5872–5881 (2017)
67. Liu, Y., Zhang, D., Lu, G., Ma, W.Y.: A survey of content-based image retrieval with high-level semantics. Pattern Recognition **40**(1), 262–282 (2007)
68. Lotfabadi, M.S., Shiratuddin, M.F., Wong, K.W.: Utilising Fuzzy Rough Set Based on Mutual Information Decreasing Method for Feature Reduction in an Image Retrieval System. In: Innovations and Advances in Computing, Informatics, Systems Sciences, Networking and Engineering, pp. 177–184 (2015)
69. Loughrey, J., Cunningham, P.: Overfitting in wrapper-based feature subset selection: The harder you try the worse it gets. In: Research and Development in Intelligent Systems XXI, pp. 33–43 (2005)
70. Lu, D., Weng, Q.: A survey of image classification methods and techniques for improving classification performance. International Journal of Remote Sensing **28**(5), 823–870 (2007)
71. Lu, J., Zhao, T., Zhang, Y.: Feature selection based-on genetic algorithm for image annotation. Knowledge-Based Systems **21**(8), 887–891 (2008)
72. Ma, L., Li, M., Gao, Y., Chen, T., Ma, X., Qu, L.: A novel wrapper approach for feature selection in object-based image classification using polygon-based cross-validation. IEEE Geoscience and Remote Sensing Letters **14**(3), 409–413 (2017)
73. Ma, Z., Nie, F., Yang, Y., Uijlings, J.R., Sebe, N.: Web image annotation via subspace-sparsity collaborated feature selection. IEEE Transactions on Multimedia **14**(4), 1021–1030 (2012)
74. Maaten, L.v.d., Hinton, G.: Visualizing data using t-SNE. Journal of Machine Learning Research **9**, 2579–2605 (2008)

75. Makadia, A., Pavlovic, V., Kumar, S.: A new baseline for image annotation. In: European Conference on Computer Vision, pp. 316–329 (2008)
76. Mui, J.K., Fu, K.S.: Automated classification of nucleated blood cells using a binary tree classifier. IEEE Transactions on Pattern Analysis and Machine Intelligence **2**(5), 429–443 (1980)
77. Ng, W.W., Dorado, A., Yeung, D.S., Pedrycz, W., Izquierdo, E.: Image classification with the use of radial basis function neural networks and the minimization of the localized generalization error. Pattern Recognition **40**(1), 19–32 (2007)
78. Núñez, J., Llacer, J.: Astronomical image segmentation by self-organizing neural networks and wavelets. Neural Networks **16**(3), 411–417 (2003)
79. Pal, M., Foody, G.M.: Feature selection for classification of hyperspectral data by SVM. IEEE Transactions on Geoscience and Remote Sensing **48**(5), 2297–2307 (2010)
80. Patil, U., Mudengudi, U.: Image fusion using hierarchical PCA. In: International Conference on image Information Processing, pp. 1–6 (2011)
81. Perez, C.A., Estévez, P.A., Vera, P.A., Castillo, L.E., Aravena, C.M., Schulz, D.A., Medina, L.E.: Ore grade estimation by feature selection and voting using boundary detection in digital image analysis. International Journal of Mineral Processing **101**(1), 28–36 (2011)
82. Picard, R.W., Minka, T.P.: Vision texture for annotation. Multimedia systems **3**(1), 3–14 (1995)
83. Porebski, A., Vandenbroucke, N., Macaire, L.: Comparison of feature selection schemes for color texture classification. In: International Conference on Image Processing Theory Tools and Applications, pp. 32–37 (2010)
84. Qi, C., Zhou, Z., Sun, Y., Song, H., Hu, L., Wang, Q.: Feature selection and multiple kernel boosting framework based on PSO with mutation mechanism for hyperspectral classification. Neurocomputing **220**, 181–190 (2017)
85. Raut, S.A., Raghuwanshi, M., Dharaskar, R., Raut, A.: Image segmentation–a state-of-art survey for prediction. In: International Conference on Advanced Computer Control, pp. 420–424 (2009)
86. Remeseiro, B., Bolon-Canedo, V., Peteiro-Barral, D., Alonso-Betanzos, A., Guijarro-Berdinas, B., Mosquera, A., Penedo, M.G., Sanchez-Marono, N.: A methodology for improving tear film lipid layer classification. IEEE Journal of Biomedical and Health Informatics **18**(4), 1485–1493 (2014)
87. Remeseiro, B., Penas, M., Barreira, N., Mosquera, A., Novo, J., García-Resúa, C.: Automatic classification of the interferential tear film lipid layer using colour texture analysis. Computer methods and programs in biomedicine **111**(1), 93–103 (2013)
88. Roffo, G., Melzi, S., Cristani, M.: Infinite feature selection. In: IEEE International Conference on Computer Vision, pp. 4202–4210 (2015)
89. Russell, B.C., Torralba, A., Murphy, K.P., Freeman, W.T.: LabelMe: a database and web-based tool for image annotation. International Journal of Computer Vision **77**(1), 157–173 (2008)
90. Sankaran, A., Jain, A., Vashisth, T., Vatsa, M., Singh, R.: Adaptive latent fingerprint segmentation using feature selection and random decision forest classification. Information Fusion **34**, 1–15 (2017)
91. Schreiber, A.T., Dubbeldam, B., Wielemaker, J., Wielinga, B.: Ontology-based photo annotation. IEEE Intelligent Systems **3**, 66–74 (2001)
92. Schroff, F., Criminisi, A., Zisserman, A.: Object Class Segmentation using Random Forests. In: British Machine Vision Conference, pp. 1–10 (2008)
93. Shafarenko, L., Petrou, M., Kittler, J.: Automatic watershed segmentation of randomly textured color images. IEEE Transactions on Image Processing **6**(11), 1530–1544 (1997)
94. Shang, C., Barnes, D.: Fuzzy-rough feature selection aided support vector machines for mars image classification. Computer Vision and Image Understanding **117**(3), 202–213 (2013)
95. Shang, C., Barnes, D., Shen, Q.: Facilitating efficient mars terrain image classification with fuzzy-rough feature selection. International Journal of Hybrid Intelligent Systems **8**(1), 3–13 (2011)
96. Shen, L., Zhu, Z., Jia, S., Zhu, J., Sun, Y.: Discriminative Gabor feature selection for hyperspectral image classification. IEEE Geoscience and Remote Sensing Letters **10**(1), 29–33 (2013)
97. Shi, C., Ruan, Q., Guo, S., Tian, Y.: Sparse feature selection based on L 2, 1/2-matrix norm for web image annotation. Neurocomputing **151**, 424–433 (2015)

98. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)
99. Thomaz, C.E., Giraldi, G.A.: A new ranking method for principal components analysis and its application to face image analysis. Image and Vision Computing **28**(6), 902–913 (2010)
100. Tuia, D., Camps-Valls, G., Matasci, G., Kanevski, M.: Learning relevant image features with multiple-kernel classification. IEEE Transactions on Geoscience and Remote Sensing **48**(10), 3780–3791 (2010)
101. Vergara, J.R., Estévez, P.A.: A review of feature selection methods based on mutual information. Neural computing and applications **24**(1), 175–186 (2014)
102. Wang, K., He, R., Wang, L., Wang, W., Tan, T.: Joint feature selection and subspace learning for cross-modal retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence **38**(10), 2010–2023 (2016)
103. Wang, Z., Feng, Y., Qi, T., Yang, X., Zhang, J.J.: Adaptive multi-view feature selection for human motion retrieval. Signal Processing **120**, 691–701 (2016)
104. Weinberger, K.Q., Saul, L.K.: Unsupervised learning of image manifolds by semidefinite programming. International Journal of Computer Vision **70**(1), 77–90 (2006)
105. Wen, X., Shao, L., Fang, W., Xue, Y.: Efficient feature selection and classification for vehicle detection. IEEE Transactions on Circuits and Systems for Video Technology **25**(3), 508–517 (2015)
106. Xue, B., Zhang, M., Browne, W., Yao, X.: A Survey on Evolutionary Computation Approaches to Feature Selection. IEEE Transactions on Evolutionary Computation **20**(4), 606–626 (2016)
107. Yao, C., Han, J., Nie, F., Xiao, F., Li, X.: Local Regression and Global Information-Embedded Dimension Reduction. IEEE Transactions on Neural Networks and Learning Systems **29**(10), 4882–4893 (2018)
108. Zeng, Z., Wang, X., Chen, Y.: Multimedia annotation via semi-supervised shared-subspace feature selection. Journal of Visual Communication and Image Representation **48**, 386–395 (2017)
109. Zhang, D., Islam, M.M., Lu, G.: A review on automatic image annotation techniques. Pattern Recognition **45**(1), 346–362 (2012)
110. Zhang, H., Fritts, J.E., Goldman, S.A.: Image segmentation evaluation: A survey of unsupervised methods. Computer Vision and Image Understanding **110**(2), 260–280 (2008)
111. Zhang, R., Nie, F., Li, X.: Self-weighted supervised discriminative feature selection. IEEE Transactions on Neural Networks and Learning systems **29**(8), 3913–3918 (2018)
112. Zhang, X., Wang, W., Li, Y., Jiao, L.: PSO-based automatic relevance determination and feature selection system for hyperspectral image classification. Electronics Letters **48**(20), 1263–1265 (2012)
113. Zhao, W., Du, S.: Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach. IEEE Transactions on Geoscience and Remote Sensing **54**(8), 4544–4554 (2016)
114. Zhao, Z.A., Liu, H.: Spectral feature selection for data mining. CRC Press (2011)
115. Zheng, W., Zhu, X., Zhu, Y., Zhang, S.: Robust Feature Selection on Incomplete Data. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, pp. 3191–3197 (2018)
116. Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., Oliva, A.: Learning deep features for scene recognition using places database. In: Advances in Neural Information Processing Systems, pp. 487–495 (2014)
117. Zhou, X., Gao, X., Wang, J., Yu, H., Wang, Z., Chi, Z.: Eye Tracking Data Guided Feature Selection for Image Classification. Pattern Recognition **63**, 56–70 (2017)
118. Zhu, C., Jia, H., Lu, T., Tao, L., Song, J., Xiang, G., Li, Y., Xie, X.: Adaptive feature selection based on local descriptor distinctive degree for vehicle retrieval application. In: IEEE International Conference on Consumer Electronics, pp. 66–69 (2017)
119. Zou, Q., Ni, L., Zhang, T., Wang, Q.: Deep learning based feature selection for remote sensing scene classification. IEEE Geoscience and Remote Sensing Letters **12**(11), 2321–2325 (2015)