

How automated image analysis techniques help scientists in species identification and classification?

E. Yousef Kalafi¹, C. Town², S. Kaur Dhillon¹

¹*Institute of Biological Sciences, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia*

²*Computer Laboratory, University of Cambridge, United Kingdom*

[Received: 30 June 2017; Accepted: 27 July 2017]

Identification of taxonomy at a specific level is time consuming and reliant upon expert ecologists. Hence the demand for automated species identification increased over the last two decades. Automation of data classification is primarily focussed on images while incorporating and analysing image data has recently become easier due to developments in computational technology. Research efforts on identification of species include specimens' image processing, extraction of identical features, followed by classifying them into correct categories. In this paper, we discuss recent automated species identification systems, mainly for categorising and evaluating their methods. We reviewed and compared different methods in step by step scheme of automated identification and classification systems of species images. The selection of methods is influenced by many variables such as level of classification, number of training data and complexity of images. The aim of writing this paper is to provide researchers and scientists an extensive background study on work related to automated species identification, focusing on pattern recognition techniques in building such systems for biodiversity studies. (Folia Morphol 2018; 77, 2: 179–193)

Key words: automated image recognition, digital image processing, species images, species classification, life data technology

INTRODUCTION

Manual identification of species according to correct taxonomy is generally impeded by obstacles such as declining the number of taxonomists and the increase in the number of described species, also discussed in detail by Gaston and O'Neill [28] and Soberon and Peterson [95], which makes identification of specimens to species a difficult and time consuming task. Automated tools may significantly assist in species recognition by facilitating reliable identification of any specimens in a population. Automated methods

that rely on pattern recognition and image analysis have been widely applied for recognition and categorisation of biological images in the field of biodiversity [27, 45, 55, 61, 68, 85, 98, 103, 108, 111]. Content based retrieval is one of the common text-based approaches in image retrieval domain [12, 94, 110] in a way that images of specimens are matched with images in database according to visual content (colour, shape, texture) similarities. For identification of species, visual features that are extracted from images based on morphology and taxonomical information play an essential role.

Address for correspondence: Associate Prof. Dr. Sarinder Kaur A/p Kashmir Singh, Institute of Biological Sciences, Faculty of Science, University of Malaya, 50603 Kuala Lumpur, Malaysia, tel: +603-79676741, fax: +603-79674178, e-mail: sarinder@um.edu.my

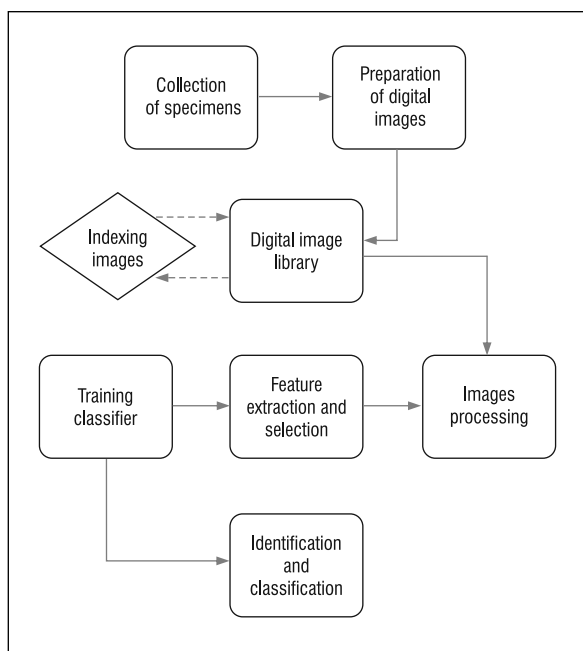


Figure 1. Schema of process for development of basic automated identification system for species images.

Taxonomic characters have been used as diagnostic features to train systems for classification of species. It is a daunting task to define the boundaries between species due to several factors such as limited morphologic features, variations among species and species' overlapping in morphometric space [5, 48], however methods that can precisely discriminate similar species according to their morphometric measurements are available.

Automated identification methods increase the possibility of extracting and analysing informative features of species in images. The image processing and analysis systems have been employed for recognition of many biological organisms [40, 77, 91, 93, 106].

Figure 1 shows the generic pipeline in building image recognition systems. Initially, biological specimens are collected and digitised. The digitising equipment alternates based on biological specimens. Microorganisms require special facility to prepare a digital image such as digital cameras, attached to microscope. All the digital images of specimens are then indexed according to their taxonomic rank or classification to prepare a digital library, often called a database. These images go through image processing using relevant techniques to extract useful features that can be used in classification. At this stage, by using data mining techniques principal component

analysis [43], projection pursuit [41], correlation existing among features [107], linear discriminant analysis [54], a huge feature vector space can be reduced. This exercise is normally referred as feature selection. Finally, a classifier is trained using techniques such as Artificial Neural Network (ANN), k-nearest neighbours (KNN), Support Vector Machine (SVM) and etc.

This paper provides a general overview of methods that can be employed in automated identification systems for biological species and is arranged as follows: Section 2 discusses previous automated systems that have been implemented while describing their techniques and methods. Section 3 discusses gaps and possible further investigations and last section concludes the paper.

WHAT HAS BEEN DONE?

Environmental monitoring based on correct identification of specimens and according to their correct species or groups is an essential and cost effective task [53]. The demand for recognition of species has significantly influenced biologists to increase the facilities and proper supply of skills for identification and classification through huge databases of data. In addition, in some cases identification of species group is limited to available human domain experts [3]. Although there was undeniable potential, the development of automated identification systems has been hampered by some taxonomists who hesitated to embrace the computational methods of species identification [51]. The main reason that influenced developing image based identification system was eagerness of taxonomist to reduce the time consumed for analysing samples [9] and to significantly cut down the costs. Culverhouse et al. [18] believe that categorising specimens from species which have significant variations in their morphology is difficult. They showed that the returned accuracy by trained personnel and experts for discriminations and labelling specimens is expected to be in the range of 64% to 95% which is within the performance range of automated methods.

Automated classification of specimens' images to their corresponding species requires development of models and methods that are able to characterise species informative data in images and apply them for recognition. These systems should be combined with databases of images or text based information [66] which are normalized in terms of size, lighting and etc.

Selection of segmentation, feature extraction and classification techniques are dependent on identification taxonomic rank, in other word means classification in level of species require more detail information compare to family level. The aim is discovering semantic concepts from images to identifying and classifying the objects of interest. For characterisation of these objects, efficient features are required to build computational models [14]. Object curvature [82] from respective contour, morphological and geometrical measurements [45, 55] are good examples of different characterisation methods.

Previously, many systems have been developed for identification of biological species. In 1996, the Dinoflagellate Categorisation by Artificial Neural Network (DiCANN) [77] was one of early automated identification systems which was developed to categorise 23 species of dinoflagellate. Later on, forensic identification of mammals according to their single hair patterns under a microscope was investigated by Moyo et al. [70], while Yuan et al. [109] discussed the identification of rats up to the species level from images of their tracks. Automation of species identification systems proved that these tedious tasks could be accomplished more feasible and efficient while minimising sources of errors [48]. Examples of such systems are Automated Leafhopper Identification System (ALIS) [21], Digital Automated Identification System (DAISY) [73], Automatic Identification and characterisation of Microbial Populations (AIMS) [44], Automated Bee Identification System (ABIS) [6], Bug-Visux [32], automated identification of bacteria using statistical methods [97], an automated identification system which estimates whiteflies, aphids and thrips densities in a greenhouse [15], Species Identification Automated (SPIDA) [84], But2fly [58], Automated Insect Identification through Concatenated Histograms of Local Appearance (AIICHLA) [53], an automated identification system for algae [17], automatic recognition of biological particles in microscopic images [81], automatic species identification of live moths [67] automated image-based phenotypic analysis in zebrafish embryos [100], automatic recognition system for some cyanobacteria using image processing techniques and ANN approach [65], automatic detection of malaria parasites for estimating parasitaemia [89], automated weed classification with local pattern-based texture descriptors [2], automated processing of imaging data through multi-tiered classification of biological structures illustrated using

caenorhabditis elegans [111], automated identification of copepods using digital image processing and artificial neural network [55], automatic plant species identification using sparse representation of leaf tooth features [42], automated system for malaria parasite identification [88], a software system for automated identification and retrieval of moth images based on wing attributes [24], automatic wild animal monitoring by identification of animal species in camera-trap images using very deep convolutional neural networks [29], automated identification of anastrepha fruit flies in the fraterculus group [76] and automated identification of fish species based on otolith contour, using Short-Time Fourier Transform and Discriminant Analysis (STFT-DA) [85]. Automated systems for biological species are summarised in Table 6.

FROM IMAGE ACQUISITION TO DATABASE

In the works reported, all automated systems were connected to a database of specimens' digital images that contain different number of dominant categories. Coltelli et al. [17] believe that image acquisition is one of most important steps in designing an automated system and capturing images should be well-focused with less complexity. The acquisition condition should be clearly defined and kept equal for all images, later labelled by expert taxonomists. One of the challenges faced in creating image database is lack of standard imaging condition during image acquisition. Larios et al. [53] proposed a method for automated identification of stonefly larvae that controls imaging. The imaging apparatus in this system poses and rotates the specimens under the microscope and captures images in standard and consistent conditions. In microscopic images magnifications might be different in the data set and it is important to specify scales in each image to prevent system confusion. Figure 2 illustrates three images of *Euryhaliotrema* with different acquisition problems.

Data in a database is commonly divided into two parts, one for training and the other set for testing the system. The number of species images used for training differs widely between systems and is determined according to the classifiers applied. Table 1 demonstrates some databases used in automated identification systems. Abu et al. [1] proposed an image retrieval framework for monogeneans that contains two databases, the monogenean image database and the MHBI-Fish ontologies. An ontology

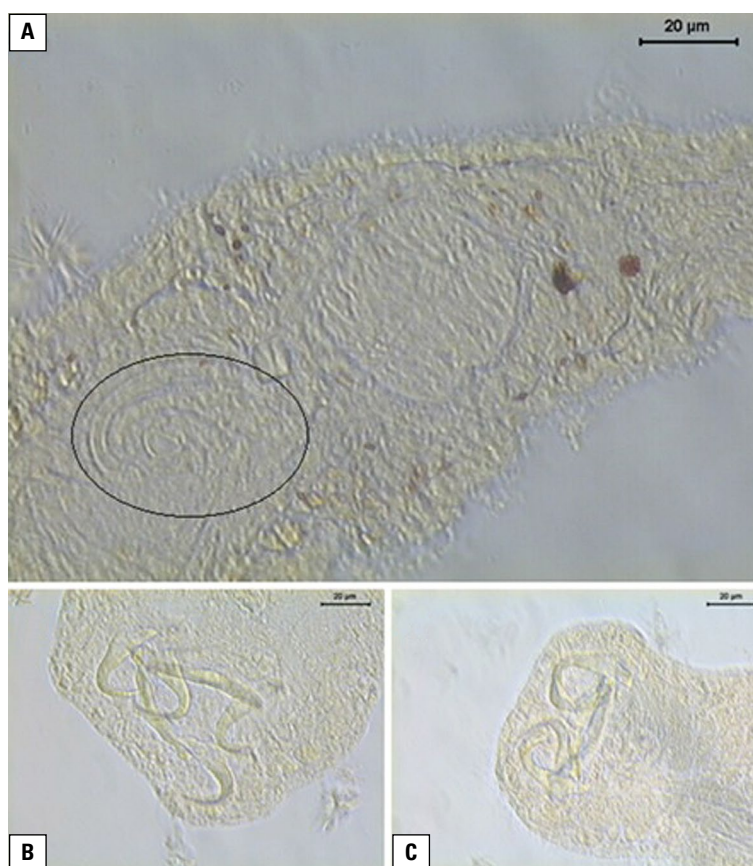


Figure 2. Illustration of image acquisition problems of *Euryhaliotrema* during digitisation. **A.** Noise and debris in the images makes recognition of copulatory organ difficult; **B.** Bad focus on bars and anchors; **C.** Messy background of anchors and bars.

Table 1. Databases for automated identification systems of species images

| Name of species | Level of classification | No. of classes | No. of training set | No. of testing data | Total no. of images | Ref. |
|-----------------|-------------------------|----------------|---------------------|---------------------|---------------------|------|
| Monogenean | Species | 6 | 148 | 19 | 167 | [1] |
| Stonefly | Species | 4 | 50 of each spp | 50 of each spp | 1240 | [53] |
| Diatom | Species | 43 | – | – | 901 | [40] |
| Forest species | Species | 112 | 8 of each spp | 8 of each spp | 2240 | [66] |
| Copepods | Genus | 5 | 30 of each spp | 20 of each spp | 400 | [68] |

framework improves the relevancy of the training set to collect the most relevant images to be used. In the stonefly identification system [53], 263 specimens of 4 species were collected and approximately 10 images of each specimen were captured through their imaging apparatus. The database used in the diatom identification system [40] includes two sets of files, the first consists of 120 images of 6 species from one genera and the second set contains 781 images of 37 species from different genera. The microscopic images in this system varied in quality and noise of con-

tour (contours being noisy) but the system was able to deal with the noise. An automated identification system for classification of tree species [66] employed a database of 112 species' images. The microscopic images were acquired with 100× magnification and labelled by dendrologists. The database contained 2240 images, 20 images from each species for training and testing the system. They used 40% of their data for training (8 images for each species), 20% for validation (4 images for each species) and 40% for testing (4 images for each species).

Role of image processing

The aim of image processing in the system is to transform digital images to a standard pose [30] and achieving recognisable objects on a uniform background. In order to facilitate the segmentation step, image artefacts should be removed, also contrast and dynamic range have to be improved. Image enhancement can be carried out by manual or automatic methods. Manual methods such as the ones carried out using ImageJ [51, 67] or Photoshop [53], may yield better image pre-processing results. It is advisable to use fully automated methods to build systems with large number of images as the manual image processing methods require longer processing time.

Digital images of species, especially microscopic images, usually contain dust or other noise artefacts. Noise makes neighbouring pixel values clutter [97], so it should be reduced by smoothing methods of filtering. It is important to know the prevalent types of noise to be filtered so that it can be removed more efficiently. Amplifier or Gaussian, salt and pepper, film grain, non-isotropic, speckle and periodic noise are the most common types of noise. Noise reduction filters can be divided into two categories: linear filters and non-linear filters [71]. Median filtering [10] is a non-linear filtering which is commonly applied to digital microscopic image [8, 35, 87, 102]. Leow et al. [55] applied median filtering with 10×10 kernel in automated identification system for copepods to suppress the salt and pepper noise created from the water in images.

Image quality is highly affected by illumination, contrast, focus and acquisition resolution [14]. Variation in illumination may also cause by different types of lenses [7] and light sources [11, 87]. Histogram equalisation can be applied to reduce variation in illumination [14]. Enhancing contrast by stretching the histogram of digital image will spread the brightest and darkest pixel values of grey levels which will later assign to white and black. Table 2 illustrates some image processing algorithms, introduced by Gonzales and Wood [30].

One of the fundamental stages before feature extraction and classification is segmentation [34]. Segmentation separates the background from the foreground and is important in computer vision since it finds the location of pixels that can be classified as an object. Pixels with common characteristics (for example texture or colour distribution) are grouped

Table 2. Image processing algorithms used in automated species identification systems. Ref. [30]

| Algorithm | Comments |
|--------------------|----------------------------------------------------|
| Noise reduction | Linear filtering, non-linear filtering |
| Image enhancement | Sharpening the image |
| | Edge highlighting |
| | Contrast improvement |
| Image restoration | Clearing away the blurriness made by linear motion |
| | Clearing away the optical misrepresentation |
| | Clearing away the periodic interference |
| Image segmentation | Separation of particular shapes from background |
| | Partitioning an image |

according to the selected segmentation algorithms. Although automated segmentation of specimens from background may still encompass debris and clutter, robust automated systems can categorise species satisfactorily [77]. Recognition of image parts which belong to an object of interest is often more effective when making use of boundaries and shape information extracted by segmentation methods. The Grabcut algorithm [83] is a segmentation technique used in automated identification of species systems [55] to remove background. In this technique, hard segmentation made by iterative graph-cut optimisation is combined with border matting to get rid of mixed and blurred pixels on boundaries of object. Edge detection [30] is another common segmentation technique that can be achieved by filters such as Canny's [13] or Sobel's [30]. In the automatic algal identification system [72], due to the significant edges and contours of the objects, both Sobel and Canny detectors were applied for image segmentation. There are generally six methods for object segmentations: Thresholding [30], fuzzy theory-based, Partial Differential Equation-based, Artificial Neural Network-based, region-based and edge-based methods [46, 50].

Thresholding is the most common technique in which binary images are produced according to cut-off value. This method can be mainly subclassed to dynamic, global and local thresholding techniques (Table 3) [46, 93]. Sometimes, especially in microscopic images, there are specimens overlapping that makes object detection difficult. Distance transforms and watershed transforms can be applied to separate overlapping specimens [20, 89].

Table 3. Thresholding techniques used in automated species identification systems

| Techniques | Subclasses | Ref. |
|------------|------------------------|------|
| Dynamic | Watershed thresholding | [22] |
| Global | Otsu thresholding | [88] |
| Local | Adaptive thresholding | [42] |

Extracting and selecting features

Features extracted from digital images are used in classification techniques; therefore, selection of best features is pivotal. Sets of features can be grouped into feature vectors which constitute a representation of objects of interest and should encompass taxonomic information. The quantity of features which are present an object of interest in image might be huge and adopting all of features into classifier will cause heavy computational effort, therefore, parsimonious selection of effective features is an important task [86]. By using feature selection techniques, the number of features that are selected for training the classifiers will be optimised [16]. Good performance of both extracted and selected features depends on type of system's classifiers and the analysing data [51]. If employed classifiers are strong enough, even if some features may be left undetected, the method may yield successful results [53].

Feature extraction

The most salient types of features in images are shape, colour and texture [37, 78, 92]. Feature extraction in automated systems may depend on the level of identification, which means features for detection at the order level are different from those at the species level. Some local features such as sparse coding spatial pyramid matching [63], concatenated feature histogram [53] and bag of words [103] which are based on scale-invariant feature transform (SIFT) [62, 101] may not collect enough information for identification of high level categories like species. In the automated system for whiteflies, aphids and thrips identification, features such as size, shape of boundary and colour components were considered [15], due to different attached part of each insect, morphological boundary was not used and only three colour components and size were applied as feature. Figure 3 illustrates some features for extraction [57]. Shape representation techniques (Table 4) [105] are applicable for shape feature extraction. The tech-

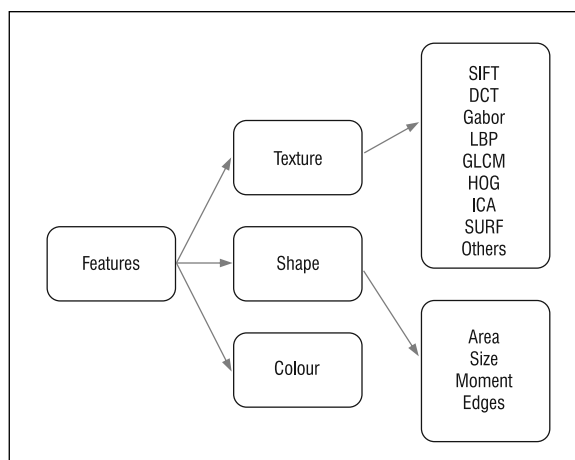


Figure 3. Content based features; SIFT — scale-invariant feature transform; DCT — discrete cosine transform; LBP — local binary patterns; GLCM — grey-level co-occurrence matrix; HOG — histogram of oriented gradient; ICA — independent component analysis; SURF — speeded-up robust features.

niques in Table 4 are classified by their processing approaches.

In the automated system for malaria parasites [88], area, perimeter, minor and major axis of red blood cells (RBC) were calculated as shape feature components. Texture features consist of kurtosis, momentum, standard deviation and mean of RBC and intensity values of the green channel were considered as colour features. Local binary patterns (LBP) [74] were considered as texture descriptors and they are applied in images analysis. Kaya et al. [49] extracted five texture features: average, correlation, entropy and energy from the LBP matrix in their automated identification system for butterfly species. In the automated identification and classification system for algae [17], dissimilarity measurement, centroid distance spectrum, points of contours and some densitometry and morphological features like area, ferret diameters, extinction, centre of gravity coordinates, etc. were calculated. Hernandez-Serna et al. [35] proposed an automated system which is applicable to identification and classification of plants, fishes and butterflies. Their approach in this system extended to three different taxonomic groups, therefore, extraction of features should be as general as possible in the way that it could be applied to all species. They used area, perimeter, diameter, compatibility, compactness and solidity as geometrical features, uniformity, median, entropy, variance, inertia, homogeneity and co-occurrence as texture features and Hu1 [69] and Ami1-Ami2 [26] as morphological features.

Table 4. Overview of shape representation techniques

| Shape features | | |
|---------------------------------------------------|------------------------------------|-------------------------------------|
| Shape parameters | Centre of gravity | |
| | Axis of least inertia | |
| | Average bending energy | |
| | Eccentricity | Principal axes method |
| | | Minimum bounding rectangle |
| | Circularity ratio | |
| | Ellipse variance | |
| | Rectangularity | |
| | Convexity | |
| | Solidity | |
| One dimensional function for shape representation | Euler number | |
| | Profiles | |
| | Hole area ratio | |
| | Complex coordinates | |
| | Centroid distance function | |
| | Tangent angle | |
| | Contour curvature | |
| | Area function | |
| | Triangle-area representation | |
| | Chord length function | |
| Polygonal approximation | Merging methods | Distance threshold method |
| | | Tunnelling method |
| | | Polygon evolution |
| Spatial interrelation feature | Splitting methods | |
| | Adaptive grid resolution | |
| | Bounding box | |
| | Convex hull | |
| | Chain code | Basic chain code |
| | | Differential chain codes |
| | | Re-sampling chain codes |
| | | Vertex chain code |
| | | Chain code histogram |
| | Smooth curve decomposition | |
| Moments | ALI-based representation | |
| | Beam angle statistics | |
| | Shape matrix | Square model shape matrix |
| | | Polar model shape matrix |
| | Shape context | |
| | Chord distribution | |
| | Shock graphs | |
| | Boundary moments | |
| | Region moments | Invariant moments |
| | | Algebraic moment invariants |
| Scale-space methods | | Zernike moments |
| | | Radial Chebyshev moments |
| | | Homocentric polar-radius moments |
| Shape transform domains | | Orthogonal Fourier-Mellin moments |
| | | Pseudo-Zernike moments |
| | Curvature scale space | |
| | Intersection points map | |
| | Fourier descriptors | One-dimensional Fourier descriptors |
| | | Region-based Fourier descriptor |
| | Wavelet transform | |
| | Angular radial transformation | |
| | Shape signature harmonic embedding | |
| | R-Transform | |
| | Shapelets descriptor | |

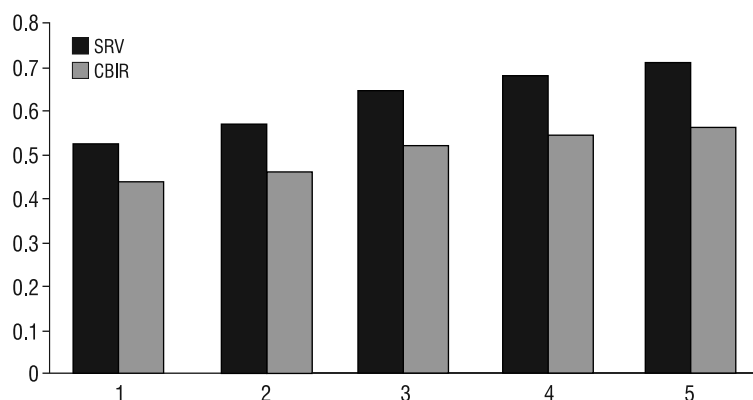


Figure 4. The comparison of mean accuracy of semantically related visual (SRV) and content based image retrieval (CBIR) approaches in categorisation of species in five iterations.

Feng and Bhanu [25] developed a system which adopted semantically related visual (SRV) attributes. They claimed that shape, texture and colour may fail in validity if the images are visually complex and have semantic contents. Figure 4 illustrates the comparison of mean accuracy of SRV and content based image retrieval (CBIR) approaches in categorisation of species in five iterations. According to the results of their research, it is notable that in all iterations accuracy of using SRV is higher than CBIR.

Other features that have been applied in detection and categorisation of specimens [51] are classical features such as branch length similarity entropy [36], corner based features, edge, ridge, curve, shape descriptors like Fourier descriptors, texture features like co-occurrence [33], histogram intensity and gradient. Also some other feature extraction methods that can be named are Gabor packet based methods [31], Histogram of Oriented Gradient [19], SIFT, Active Shape Model (ASM) [4], Active Appearance Model (AAM) [80] and LBP.

Feature selection

Feature selection is a process to identify relevant features while removing irrelevant and redundant features. Relevant features should be informative, fast in computing and also invariant to noise or given transformations. Now, the decision whether a feature is relevant, redundant or not, are aspects that involves in feature selection operations. The role of selecting features lies in improving the prediction process, correlation coefficient of regression algorithms and

comprehensibility of learning results [47]. Table 5 demonstrates some feature selection algorithms [52]. Principal component analysis [43] is multivariate statistical technique, adopted by DAISY to achieve most of important features of images. Due to big amount of detailed information collected by this technique, acquired features are convenient for identification at species level [101]. Ali et al. [3] used the assessment of Sequential Backward Selection (SBS), Sequential Forward Selection (SFS) and Sequential Forward Floating Selection (SFFS) techniques [99] for selecting proper features for monogenean classification and the results indicated that of the 25 features, 21 were the best in classification of *Gyrodactylus* species performances. Feature selection results are dependent on the size of the training data as in [38], the quality of feature selection for small data is low and as the training size increases, the quality improves.

Classification

The idea of classification is to classify objects of interest based on a specific feature data set to discriminate between distinct classes. Performance of classifiers is highly affected by the segmentation and feature extraction process. Jain et al. [39] proposed three categories of classifiers: similarity based, probabilistic and decision boundaries. Most of the classification methods are mentioned in [60, 89, 112], including structural, fuzzy, transform, neural network-based methods and many more. Some automated identification systems [55] employ neural networks or learning algorithms when there are many classes and small

Table 5. Some of feature selection algorithms used in automated species identification systems

| Algorithm | Subset | Search type |
|-------------------------------------------|---------------------------------------------------------|-------------|
| SFS, SBS | Looking for the best subset of given size | Sequential |
| GSFS(g), GSBS(g) | Looking for the best subset of given size | Sequential |
| PTA(l, r) | Looking for the best subset of given size | Sequential |
| GPTA(l, r) | Looking for the best subset of given size | Sequential |
| SFFS, SBFS | Looking for the best subset of given size | Sequential |
| BAB, BAB ⁺ , BAB ⁺⁺ | Looking for the best subset of given size | Sequential |
| RBAB, RBABM | Looking for the smallest acceptable subset | Sequential |
| GA | Looking for optimal combined size and error rate subset | Parallel |
| PARA | Looking for optimal combined size and error rate subset | Parallel |

Table 6. Examples of some automated species identification systems

| System | No. of classes | Classification method | Accuracy (%) | Ref. |
|---------------------------------------------------------------------------------------|----------------|------------------------------------------------------------------------------------|--------------|-------|
| Automated object recognition of blue-green algae | 9 | Discriminant analysis | 98 | [96] |
| Automatic classification of field-collected dinoflagellates | 23 | ANN: RBF and BP of error variant | 83 | [77] |
| Automatic identification of human helminth eggs | 12 | ANN | 86–90 | [106] |
| Automate identification of bees | 13 | Linear discriminant analysis | 98–99.8 | [91] |
| Automatic diatom identification | 43 | Decision trees and KNN | 82–84 | [40] |
| Automatic identification of whiteflies, aphids and thrips | 50 | ANN | 93–100 | [15] |
| Automatic identification of live moths | 35 | WEKA: Naïve Bayes, instance-based learning, decision trees, random forests and SVM | 85 | [67] |
| Automatic recognition system for some cyanobacteria | 4 | ANN | 95 | [65] |
| Automated weed classification | 2 | Template matching and SVM | 88–98.5 | [2] |
| Automated insect identification | 4 | Kadir entropy detector and PCBR | 82–95 | [53] |
| Automated taxon identification of teleost fishes | 420 | KNN | 72 | [75] |
| Automated real-time dynamic identification of flying and resting butterfly | 10 | Random tree | 85 | [59] |
| Automatic identification of diatoms | 12 | BP neural networks | 94 | [64] |
| Automatic insect classification | 10 | SVM | > 90 | [56] |
| Automated identification and retrieval of moth images | 50 | SRV | 85 | [25] |
| Automatic identification of species | 740 | ANN | 91–93 | [35] |
| Water monitoring — automated and real time identification and classification of algae | 23 | ANN: SOM | 98 | [17] |
| Automatic identification of butterfly species | 5 | ANN | 98 | [49] |
| Automated system for malaria parasite identification | 2 | SVM | 80 | [88] |
| Automatic plant species identification | 8 | Sparse representation | 76–79 | [42] |
| Automated identification of copepods | 8 | ANN | 93.13 | [55] |
| Automated identification and retrieval of moth images | 50 | SRV attributes | 34–70 | [24] |
| Automatic wild animal identification | 26 | Convolutional neural networks | 88.9–98.1 | [29] |

ANN — artificial neural network; BP — back propagation; KNN — k-nearest neighbours; PCBR — principal curvature-based region; RBF — radial basis function; SOM — self organising map; SRV — semantically-related visual; SVM — support vector machine

number of samples, but some other systems [75] that deal with huge numbers of samples preferably use other algorithms like KNN [23]. As a summary, Table 6 indicates some automated identification systems adopting various kinds of classification methods.

Jalba et al. [40] used KNN and C4.5 [79] algorithms as classification techniques for an automated identification system. In this system two types of feature vectors were adopted. Both types of feature vectors were constructed for top and bottom curvature spaces. Type 1 feature vector computes the number of peaks, mean curvature and variance for each cluster. Type 2 feature vector computes the mean curvature and variance of the points with the highest curvature for each cluster and the extent. The result with type 2 feature vectors was 84% and better than type 1 feature vectors. Comparing the rate of identification with human experts (43% to 86.5%), the average accuracy of this system when using C4.5 decision trees was higher. Mayo and Watson [67] employed methods from the WEKA [104] machine learning toolkit, including Naïve Bayes, J48, IB1, IB5, Random forests and sequential minimal optimisation (SMO) classifiers. The results demonstrated that random forest and SMO classifiers achieved accuracy of 83%, better than other classifiers and by increasing the number of feature attributes, the accuracy reaches to 85%. In identification of species of *Gyrodactylus* genus in fish ectoparasite [4], features which were extracted by active shape models (ASM), implemented to two linear classifiers, linear discriminant analysis (LDA) and KNN and two non-linear classifiers, multilayer perceptron (MLP) and support vector machine (SVM). According to results of the study, LDA method accuracy was 85.71%, MLP method 95.59% and KNN classification accuracy of 98.75%. KNN was outperforming classifier since the testing dataset in their study is 68 images and KNN is capable of classifying with limited number of dataset. Mansoor et al. [65] proposed a system operating with ANN for identification of cyanobacteria genera images. This system recognised 71 of 80 images correctly and detection accuracy was reported as being 95%. Le-Qing and Zhen [56] employed two SVM classifiers using radial basis functions (RBF) and polynomial kernels respectively in their automated insect identification system. Comparing the evaluated results of these two classifiers, it is notable that polynomial kernel performs better than RBF in verification (91.96–87.5%) and RBF performs better than polynomial kernel in discrimination (93.35–91.57%).

These two SVM classifiers were also employed in an automated classification system for Erythrocytes infected with malaria [90]. With combination of both classifiers, an identification accuracy rate of 96.42% was achieved. In automated identification of insects at the order level [101], ANN and SVM were used as classification methods. Since SVM is a binary classifier and for classification of multi-class problem it has to use one over all classification for each class, SVM performs better than ANN. Comparing SVM and ANN results with SRV attributes in an automated identification system for moths [25], this classifier outperforms both SVM and ANN classifiers. In [49], classification was based on LBP and the accuracy rate in identification depends on variables such as neighbouring and radius values.

FUTURE DIRECTIONS

Automated identification systems can help biologists to identify and classify species with respect to time and human energy efficiency. Much work has been done in this field yet there are gaps and challenges that may affect the performances of systems such as lower accuracy of classification and identification of species compared to human experts. Using systems may not affect the number of people who involve in the process. Still the specimens' collection should be handled by human and system's performance has to be monitored by taxonomist experts.

Improvements in the technology and systems can minimise problems, issues and constraints in manual categorisation. An important challenge in species classification is the lack of a consistent database of digital images. Most of the systems reported in this paper contain small datasets of species from few classes and rarely representative databases were used. To the best of our knowledge, proposed methods for normalisation of illumination variations, especially in complex images, can be improved. Automated segmentation of complex images is still an open area for investigation, specifically for micro-organisms which are stained on slides and recognition of their background and foreground is difficult. In most of the discussed systems, images of species were taken on a planar and uniform background, but the need to work with non-uniform or cluttered images is unavoidable thus automated segmentation methods must be improved for this case. Mostly, systems are limited to two-dimensional images of specimens

whereas developing methods that deal with three-dimensional images may reliably accelerate the identification.

Although many automated systems have been developed and implemented such as those mentioned in this paper, there are only a limited number that are accessible to end users like DAISY [73] and SPIDA-Web [84]. Biologists prefer to keep their specimens' samples in their labs and hesitate to employ such systems while digitising specimens optimises conservation space and economically is effective.

Considering the lack of computational systems in biodiversity still there is a trust gap between biologists and computational experts. Lack of attention to biodiversity informatics hampered the growth of biodiversity systems whereas these systems can be a remedy for many biomedical topics.

CONCLUSIONS

Over the last two decades several systems have been designed and developed for automatic identification of species. The demand for accurate recognition of biological objects up to species level caused an incredible interest in implementation of new identification systems. This paper aimed to summarise step by step the techniques (Fig. 1) used in existing automated identification systems.

Generally, the operation accuracy of automated identification systems for biological images is highly dependent on quality of acquired digital images. The significant challenge is converting complex images into simple ones. Still, if the quality of images are not convincing, image processing techniques are powerful assistance to this conversion. Each step in building automated identification systems for species images should be compatible with next step. The image processing stage has to highlight those information in images which will be extracted for classification.

In practice, with regards to image resolution, magnification and illumination, the accuracy rate of each system may differ. Also the features extracted from images directly influence the operation of classifiers. Although automated recognition systems help researchers in building species taxa, none of the automated systems work with consistent accuracy under different conditions and this uncertainty about their performance gives rise to a continuing need for verification by taxonomists and experts.

Acknowledgements

Authors are thankful to the late Susan Lim Lee Hong and Tan Wooi Boon for tendering their valuable experiences in biological laboratory.

Funding

This project was supported by University of Malaya Postgraduate Research Fund (PG092-2013B) to the first author and the University of Malaya Research Grant (UMRG) Programme Based Grant (RP008-2012A) and the University of Malaya's Living Lab Grant Programme — Sustainability Science (LL020-16SUS) to the fourth author.

REFERENCES

1. Abu A, Lim S, Sidhu A, et al. Biodiversity image retrieval framework for monogeneans. *Syst Biodivers*. 2013; 11(1): 19–33, doi: [10.1080/14772000.2012.761655](https://doi.org/10.1080/14772000.2012.761655).
2. Ahmed F, Kabir MH, Bhuyan S, et al. Automated weed classification with local pattern-based texture descriptors. *Int Arab J Inf Technol*. 2014; 11: 87–94.
3. Ali R, Hussain A, Bron J, et al. Multi-stage classification of Gyrodactylus species using machine learning and feature selection techniques. *IEEE*. 2011: 457–462, doi: [10.1109/isda.2011.6121698](https://doi.org/10.1109/isda.2011.6121698).
4. Ali R, Hussain A, Bron JE, Shinn AP: The Use of ASM Feature Extraction and Machine Learning for the Discrimination of Members of the Fish Ectoparasite Genus Gyrodactylus; in Huang T, Zeng Z, Li C, Leung CS (eds): *Neural Information Processing*. Springer Berlin Heidelberg, 2012, [cited 2016 Sep 30], pp 256–263.
5. Anton A, Duthie H. Use of cluster analysis in the systematics of the algal genus *Cryptomonas*. *Can J Bot*. 1981; 59(6): 992–1002, doi: [10.1139/b81-136](https://doi.org/10.1139/b81-136).
6. Arbuckle T, Schröder S, Steinhage V, Wittmann D: *Biodiversity Informatics in Action: Identification and Monitoring of Bee Species Using ABIS* 2001.
7. Arce SH, Wu PH, Tseng Y. Fast and accurate automated cell boundary determination for fluorescence microscopy. *Sci Rep*. 2013; 3: 2266, doi: [10.1038/srep02266](https://doi.org/10.1038/srep02266), indexed in Pubmed: [23881180](https://pubmed.ncbi.nlm.nih.gov/23881180/).
8. Avci D, Varol A. An expert diagnosis system for classification of human parasite eggs based on multi-class SVM. *Expert Syst Appl*. 2009; 36(1): 43–48, doi: [10.1016/j.eswa.2007.09.012](https://doi.org/10.1016/j.eswa.2007.09.012).
9. Benfield M, Grosjean P, Culverhouse P, et al. RAPID: research on automated plankton identification. *Oceanography*. 2007; 20(2): 172–187, doi: [10.5670/oceanog.2007.63](https://doi.org/10.5670/oceanog.2007.63).
10. Bovik AC, Huang TS, Munson DC. The effect of median filtering on edge estimation and detection. *IEEE Trans Pattern Anal Mach Intell*. 1987; 9(2): 181–194, indexed in Pubmed: [21869390](https://pubmed.ncbi.nlm.nih.gov/21869390/).
11. Bradbury MH, Bracegirdle DB. *Introduction to Light Microscopy*. 2 ed. Oxford, Garland Science. 1998.

12. Bunte K, Biehl M, Jonkman M, et al. Learning effective color features for content based image retrieval in dermatology. *Pattern Recognition*. 2011; 44(9): 1892–1902, doi: [10.1016/j.patcog.2010.10.024](https://doi.org/10.1016/j.patcog.2010.10.024).
13. Canny J. A Computational Approach to Edge Detection. *IEEE Trans Pattern Anal Mach Intell*. 1986; PAMI-8(6): 679–698, doi: [10.1109/tpami.1986.4767851](https://doi.org/10.1109/tpami.1986.4767851).
14. Castañón C, Fraga J, Fernandez S, et al. Biological shape characterization for automatic image recognition and diagnosis of protozoan parasites of the genus *Eimeria*. *Pattern Recognition*. 2007; 40(7): 1899–1910, doi: [10.1016/j.patcog.2006.12.006](https://doi.org/10.1016/j.patcog.2006.12.006).
15. Cho J, Choi J, Qiao, et al. Automatic Identification of Tobacco Whiteflies, Aphids and Thrips in Greenhouse Using Image Processing Techniques Mexico, 2008, [cited 2016 Sep 30], pp 46–53.
16. Choras RS: Image feature extraction techniques and their applications for CBIR and biometrics systems. *Int J Biol Biomed Engineering* 2007; 1: 6–16.
17. Coltelli P, Barsanti L, Evangelista V, et al. Water monitoring: automated and real time identification and classification of algae using digital microscopy. *Environ Sci Process Impacts*. 2014; 16(11): 2656–2665, doi: [10.1039/c4em00451e](https://doi.org/10.1039/c4em00451e), indexed in Pubmed: [25294420](https://pubmed.ncbi.nlm.nih.gov/25294420/).
18. Culverhouse PF, Williams R, Reguera B, et al. Do experts make mistakes? A comparison of human and machine identification of dinoflagellates. *Mar Ecol Prog Ser*. 2003; 247: 17–25, doi: [10.3354/meps247017](https://doi.org/10.3354/meps247017).
19. Dalal N, Triggs B: Histograms of Oriented Gradients for Human Detection; in : *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01*. Washington, DC, USA, IEEE Computer Society, 2005, pp 886–893.
20. Ruberto CDi, Dempster A, Khan S, et al. Automatic thresholding of infected blood images using granulometry and regional extrema. *IEEE Comput. Soc.* 2000: 441–444, doi: [10.1109/icpr.2000.903579](https://doi.org/10.1109/icpr.2000.903579).
21. Dietrich CH, Pooley CD. Automated Identification of Leafhoppers (Homoptera: Cicadellidae: Draeculacephala Ball). *Ann Entomol Soc Am*. 1994; 87(4): 412–423, doi: [10.1093/aesa/87.4.412](https://doi.org/10.1093/aesa/87.4.412).
22. Doncic A, Eser U, Atay O, et al. An algorithm to automate yeast segmentation and tracking. *PLoS One*. 2013; 8(3): e57970, doi: [10.1371/journal.pone.0057970](https://doi.org/10.1371/journal.pone.0057970), indexed in Pubmed: [23520484](https://pubmed.ncbi.nlm.nih.gov/23520484/).
23. Duda RO, Hart PE, Stork DG. *Pattern Classification*. John Wiley & Sons. 2012.
24. Feng L, Bhanu B, Heraty J. A software system for automated identification and retrieval of moth images based on wing attributes. *Pattern Recognition*. 2016; 51: 225–241, doi: [10.1016/j.patcog.2015.09.012](https://doi.org/10.1016/j.patcog.2015.09.012).
25. Feng L, Bhanu B. Automated identification and retrieval of moth images with semantically related visual attributes on the wings. *IEEE*. 2013: 2577–2581, doi: [10.1109/icip.2013.6738531](https://doi.org/10.1109/icip.2013.6738531).
26. Flusser J, Suk T. Pattern recognition by affine moment invariants. *Pattern Recognition*. 1993; 26(1): 167–174, doi: [10.1016/0031-3203\(93\)90098-h](https://doi.org/10.1016/0031-3203(93)90098-h).
27. Francoy T, Wittmann D, Drauschke M, et al. Identification of Africanized honey bees through wing morphometrics: two fast and efficient procedures. *Apidologie*. 2008; 39(5): 488–494, doi: [10.1051/apido:2008028](https://doi.org/10.1051/apido:2008028).
28. Gaston KJ, O'Neill MA. Automated species identification: why not? *Philos Trans R Soc Lond B Biol Sci*. 2004; 359(1444): 655–667, doi: [10.1098/rstb.2003.1442](https://doi.org/10.1098/rstb.2003.1442), indexed in Pubmed: [15253351](https://pubmed.ncbi.nlm.nih.gov/15253351/).
29. Gomez A, Salazar A: Towards Automatic Wild Animal Monitoring: Identification of Animal Species in Camera-trap Images using Very Deep Convolutional Neural Networks. *ArXiv Prepr ArXiv160306169* 2016. <http://arxiv.org/abs/1603.06169> (cited 2016 Oct 2).
30. Gonzalez RC, Woods RE. *Digital Image Processing*. 3 edition Upper Saddle River. N.J. Pearson. 2007.
31. Grigorescu SE, Petkov N, Kruizinga P. Comparison of texture features based on Gabor filters. *IEEE Trans Image Process*. 2002; 11(10): 1160–1167, doi: [10.1109/TIP.2002.804262](https://doi.org/10.1109/TIP.2002.804262), indexed in Pubmed: [18249688](https://pubmed.ncbi.nlm.nih.gov/18249688/).
32. Hanqing Z, Zuorui S, University YXA, (China) B, Protection) C of P: On computer-aided insect identification through math-morphology features. *Zhonggou Nongye Daxue Xuebao China* 2002. <http://agris.fao.org/agris-search/search.do?recordID=CN2003080710> (cited 2016 Sep 30).
33. Haralick R, Shanmugam K, Dinstein I. Textural Features for Image Classification. *IEEE Trans Syst Man Cybern*. 1973; SMC-3(6): 610–621, doi: [10.1109/tsmc.1973.4309314](https://doi.org/10.1109/tsmc.1973.4309314).
34. Haralick RM, Shapiro LG. *Computer and Robot Vision*. ed 1st Boston, MA, USA, Addison-Wesley Longman Publishing Co., Inc. 1992.
35. Hernández-Serna A, Jiménez-Segura LF. Automatic identification of species with neural networks. *Peer J*. 2014; 2: e563, doi: [10.7717/peerj.563](https://doi.org/10.7717/peerj.563), indexed in Pubmed: [25392749](https://pubmed.ncbi.nlm.nih.gov/25392749/).
36. Huddar SR, Gowri S, Keerthana K, et al. Novel algorithm for segmentation and automatic identification of pests on plants using image processing. *IEEE*. 2012: 1–5, doi: [10.1109/icccnt.2012.6396012](https://doi.org/10.1109/icccnt.2012.6396012).
37. Islam M, Dengsheng Z, Guojun L. A geometric method to compute directionality features for texture images. *IEEE*. 2008: 1521–1524, doi: [10.1109/icme.2008.4607736](https://doi.org/10.1109/icme.2008.4607736).
38. Jain A, Zongker D. Feature selection: evaluation, application, and small sample performance. *IEEE Trans Pattern Anal Mach Intell*. 1997; 19(2): 153–158, doi: [10.1109/34.574797](https://doi.org/10.1109/34.574797).
39. Jain AK, Duin PW. Statistical pattern recognition: a review. *IEEE Trans Pattern Anal Mach Intell*. 2000; 22(1): 4–37, doi: [10.1109/34.824819](https://doi.org/10.1109/34.824819).
40. Jalba A, Wilkinson M, Roerdink J, et al. Automatic diatom identification using contour analysis by morphological curvature scale spaces. *Mach Vis Appl*. 2005; 16(4): 217–228, doi: [10.1007/s00138-005-0175-8](https://doi.org/10.1007/s00138-005-0175-8).
41. Jimenez L, Landgrebe D. High dimensional feature reduction via projection pursuit. *ECE Tech Rep*. 1996: 103, doi: [10.1109/igarss.1994.399367](https://doi.org/10.1109/igarss.1994.399367).
42. Jin T, Hou X, Li P, et al. A novel method of automatic plant species identification using sparse representation of leaf tooth features. *PLoS One*. 2015; 10(10): e0139482, doi: [10.1371/journal.pone.0139482](https://doi.org/10.1371/journal.pone.0139482), indexed in Pubmed: [26440281](https://pubmed.ncbi.nlm.nih.gov/26440281/).
43. Jolliffe I: *Principal component analysis*. Wiley Online Library, 2002. <http://onlinelibrary.wiley.com/doi/10.1002/9781118445112.stat06472/full> (cited 2016 Oct 3).

44. Jonker R, Groben R, Tarran G, et al. Automated identification and characterisation of microbial populations using flow cytometry: the AIMS project. *Scientia Marina*. 2000; 64(2): 225–234, doi: [10.3989/scimar.2000.64n2225](https://doi.org/10.3989/scimar.2000.64n2225).
45. Kalafi EY, Tan WB, Town C, et al. Automated identification of Monogeneans using digital image processing and K-nearest neighbour approaches. *BMC Bioinformatics*. 2016; 17(Suppl 19): 511, doi: [10.1186/s12859-016-1376-z](https://doi.org/10.1186/s12859-016-1376-z), indexed in Pubmed: [28155722](https://pubmed.ncbi.nlm.nih.gov/28155722/).
46. Kang WX, Yang QQ, Liang RP. The Comparative Research on Image Segmentation Algorithms. *IEEE*. 2009: 703–707, doi: [10.1109/etcs.2009.417](https://doi.org/10.1109/etcs.2009.417).
47. Karagiannopoulos M, Anyfantis D, Kotsiantis SB, Pintelas PE: Feature selection for regression problems. In : *Proceedings of the 8th Hellenic European Research on Computer Mathematics & its Applications*, Athens, Greece. 2007. <http://www.math.upatras.gr/~dany/Downloads/hercma07.pdf> (cited 2016 Oct 5).
48. Kay JW, Shinn AP, Sommerville C. Towards an automated system for the identification of notifiable pathogens: using *Gyrodactylus salaris* as an example. *Parasitol Today*. 1999; 15(5): 201–206, indexed in Pubmed: [10322355](https://pubmed.ncbi.nlm.nih.gov/10322355/).
49. Kaya Y, Kayci L, Uyar M. Automatic identification of butterfly species based on local binary patterns and artificial neural network. *Applied Soft Comput*. 2015; 28: 132–137, doi: [10.1016/j.asoc.2014.11.046](https://doi.org/10.1016/j.asoc.2014.11.046).
50. Khan W. Image Segmentation Techniques: A Survey. *J Image Graph*. 2014: 166–170, doi: [10.12720/joig.1.4.166-170](https://doi.org/10.12720/joig.1.4.166-170).
51. Kiranyaz S, Ince T, Pulkkinen J, et al. Classification and retrieval on macroinvertebrate image databases. *Comput Biol Med*. 2011; 41(7): 463–472, doi: [10.1016/j.compbiomed.2011.04.008](https://doi.org/10.1016/j.compbiomed.2011.04.008), indexed in Pubmed: [21601841](https://pubmed.ncbi.nlm.nih.gov/21601841/).
52. Kudo M, Sklansky J. Comparison of algorithms that select features for pattern classifiers. *Pattern Recognition*. 2000; 33(1): 25–41, doi: [10.1016/s0031-3203\(99\)00041-2](https://doi.org/10.1016/s0031-3203(99)00041-2).
53. Larios N, Deng H, Zhang W, et al. Automated insect identification through concatenated histograms of local appearance features: feature vector generation and region detection for deformable objects. *Mach Vis Appl*. 2007; 19(2): 105–123, doi: [10.1007/s00138-007-0086-y](https://doi.org/10.1007/s00138-007-0086-y).
54. Lei Z, Liao S, Li SZ: Efficient feature selection for linear discriminant analysis and its application to face recognition. *IEEE*, 2012, [cited 2016 Oct 2], pp 1136–1139.
55. Leow LK, Chew LL, Chong VC, et al. Automated identification of copepods using digital image processing and artificial neural network. *BMC Bioinformatics*. 2015; 16 Suppl 18: S4, doi: [10.1186/1471-2105-16-S18-S4](https://doi.org/10.1186/1471-2105-16-S18-S4), indexed in Pubmed: [26678287](https://pubmed.ncbi.nlm.nih.gov/26678287/).
56. Le-Qing Z, Zhen Z. Automatic insect classification based on local mean colour feature and Supported Vector Machines. *Oriental Insects*. 2012; 46(3-4): 260–269, doi: [10.1080/00305316.2012.738142](https://doi.org/10.1080/00305316.2012.738142).
57. Li J, Tseng KK, Hsieh ZuYi, et al. Staining pattern classification of antinuclear autoantibodies based on block segmentation in indirect immunofluorescence images. *PLoS One*. 2014; 9(12): e113132, doi: [10.1371/journal.pone.0113132](https://doi.org/10.1371/journal.pone.0113132), indexed in Pubmed: [25474260](https://pubmed.ncbi.nlm.nih.gov/25474260/).
58. Liu F, Shen ZR, Zhang JW, et al. Automatic insect identification based on color characters. *Chin Bull Entomol*. 2008; 45: 150–153.
59. Loke KS, Egerton S, Cristofaro D, Clementson S: Automated real-time dynamic identification of flying and resting butterfly species in the natural environment. In: 2011 International Conference on Environment Science and Engineering (ICESE 2011) (Saji Baby and Richard Haynes 1 April 2011 to 3 April 2011). Institute of Electrical and Electronics Engineers, 2011, [cited 2016 Oct 1], pp. 179–183.
60. Loncaric S. A survey of shape analysis techniques. *Pattern Recognition*. 1998; 31(8): 983–1001, doi: [10.1016/s0031-2023\(97\)00122-2](https://doi.org/10.1016/s0031-2023(97)00122-2).
61. Loos A, Ernst A. An automated chimpanzee identification system using face detection and recognition. *EURASIP J Image Video Process*. 2013; 2013(1): 49, doi: [10.1186/1687-5281-2013-49](https://doi.org/10.1186/1687-5281-2013-49).
62. Lowe D. Distinctive Image Features from Scale-Invariant Keypoints. *Int J Comput Vis*. 2004; 60(2): 91–110, doi: [10.1023/b:visi.0000029664.99615.94](https://doi.org/10.1023/b:visi.0000029664.99615.94).
63. Lu An, Hou X, Lin C, et al. Insect Species Recognition using Sparse Representation. *Br Machine Vision Association*. 2010: 108.1–108.10, doi: [10.5244/c.24.108](https://doi.org/10.5244/c.24.108).
64. Luo Q, Gao Y, Luo J, et al. Automatic Identification of Diatoms with Circular Shape using Texture Analysis. *J Software*. 2011; 6(3), doi: [10.4304/jsw.6.3.428-435](https://doi.org/10.4304/jsw.6.3.428-435).
65. Mansoor H, Sorayya M, Aishah S, Mogeeb A, Mosleh A. Automatic recognition system for some cyanobacteria using image processing techniques and ANN approach; In: *Int Conf on Envir and Comp Science IPCBEE*. 2011, pp 73–78.
66. Martins J, Oliveira LS, Nisgoski S, et al. A database for automatic classification of forest species. *Mach Vis Appl*. 2012; 24(3): 567–578, doi: [10.1007/s00138-012-0417-5](https://doi.org/10.1007/s00138-012-0417-5).
67. Mayo M, Watson A. Automatic species identification of live moths. *Knowledge-Based Systems*. 2007; 20(2): 195–202, doi: [10.1016/j.knosys.2006.11.012](https://doi.org/10.1016/j.knosys.2006.11.012).
68. Mehdipour Ghazi M, Yanikoglu B, Aptoula E. Plant identification using deep neural networks via optimization of transfer learning parameters. *Neurocomputing*. 2017; 235: 228–235, doi: [10.1016/j.neucom.2017.01.018](https://doi.org/10.1016/j.neucom.2017.01.018).
69. Ming-Kuei H. Visual pattern recognition by moment invariants. *IEEE Transactions on Information Theory*. 1962; 8(2): 179–187, doi: [10.1109/tit.1962.1057692](https://doi.org/10.1109/tit.1962.1057692).
70. Moyo T, Bangay S, Foster G. The identification of mammalian species through the classification of hair patterns using image pattern recognition. *Proceedings of the 4th international conference on Computer graphics, virtual reality, visualisation and interaction in Africa - Afrigraph '06*. 2006, doi: [10.1145/1108590.1108619](https://doi.org/10.1145/1108590.1108619).
71. Mythili C, Kavitha V. Efficient technique for color image noise reduction. *Res Bull Jordan ACM*. 2011; 1: 41–44.
72. Natchimuthu S, Natchimuthu S, Chinnaraj P, et al. Automatic Identification of Algal Community from Microscopic Images. *Bioinforma Biol Insights*. 2013: 327.
73. O'Neill MA, Gauld ID, Gaston KJ, Weeks PJD: Daisy: an automated invertebrate identification system using holistic vision techniques; in : *Proceedings of the Inaugural Meeting BioNET-INTERNATIONAL Group for Computer-Aided Taxonomy (BIGCAT)*. 2000, pp 13–22.

74. Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans Pattern Anal Mach Intell.* 2002; 24(7): 971–987, doi: [10.1109/tpami.2002.1017623](https://doi.org/10.1109/tpami.2002.1017623).
75. Parisi-Baradad V, Manjabacas A, Lombarte A, et al. Automated Taxon Identification of Teleost fishes using an otolith online database: AFORO. *Fisheries Res.* 2010; 105(1): 13–20, doi: [10.1016/j.fishres.2010.02.005](https://doi.org/10.1016/j.fishres.2010.02.005).
76. Perre P, Faria FA, Jorge LR, et al. Toward an Automated Identification of Anastrepha Fruit Flies in the fraterculus group (Diptera, Tephritidae). *Neotrop Entomol.* 2016; 45(5): 554–558, doi: [10.1007/s13744-016-0403-0](https://doi.org/10.1007/s13744-016-0403-0), indexed in Pubmed: [27156899](https://pubmed.ncbi.nlm.nih.gov/27156899/).
77. Phil C, Simpson RG, Rob E, et al. Automatic classification of field-collected dinoflagellates by artificial neural network. *Mar Ecol Prog Ser.* 1996; 139: 281–287, doi: [10.3354/meps139281](https://doi.org/10.3354/meps139281).
78. Ping Tian D. A review on image feature extraction and representation techniques. *Int J Multimed Ubiquitous Eng.* 2013; 8: 385–396.
79. Quinlan JR: C4. 5: programs for machine learning. Elsevier, 2014. <https://books.google.com/books?hl=en&lr=&id=b3ujBQAAQBAJ&oi=fnd&pg=PP1&dq=C4.+5:+programs+for+machine+learning.+Morgan+Kaufmann&ots=sP8rULD6&sig=Yv1JBLnQGgoc5y8JfzcwOYvBqVw> (cited 2016 Oct 1).
80. Quivy CH, Kumazawa I. Normalization of Active Appearance Models for Fish Species Identification. *ISRN Signal Processing.* 2011; 2011: 1–16, doi: [10.5402/2011/103293](https://doi.org/10.5402/2011/103293).
81. Ranzato M, Taylor PE, House JM, et al. Automatic recognition of biological particles in microscopic images. *Pattern Recognition Lett.* 2007; 28(1): 31–39, doi: [10.1016/j.patrec.2006.06.010](https://doi.org/10.1016/j.patrec.2006.06.010).
82. Riggs LA. Curvature as a feature of pattern vision. *Science.* 1973; 181(4104): 1070–1072, indexed in Pubmed: [4726002](https://pubmed.ncbi.nlm.nih.gov/4726002/).
83. Rother C, Kolmogorov V, Blake A. Grabcut: Interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics (TOG).* 2004; 23(3): 309–314, doi: [10.1145/1015706.1015720](https://doi.org/10.1145/1015706.1015720).
84. Russell KN, Do MT, Huff JC, Platnick NI, MacLeod N. Introducing SPIDA-web: wavelets, neural networks and Internet accessibility in an image-based automated identification system. *Autom Taxon Identif Syst Theory Approaches Appl* 2007; 131–152.
85. Salimi N, Loh KH, Dhillon SK, et al. Fully-automated identification of fish species based on otolith contour: using short-time Fourier transform and discriminant analysis (STFT-DA). *PeerJ.* 2016; 4: e1664, doi: [10.7717/peerj.1664](https://doi.org/10.7717/peerj.1664), indexed in Pubmed: [26925315](https://pubmed.ncbi.nlm.nih.gov/26925315/).
86. Sang-Hee L. A Novel Approach to Shape Recognition Using Shape Outline. *J Korean Phys Soc.* 2010; 56(3(1)): 1016–1019, doi: [10.3938/jkps.56.1016](https://doi.org/10.3938/jkps.56.1016).
87. Saraswat M, Arya KV. Automated microscopic image analysis for leukocytes identification: a survey. *Micron.* 2014; 65: 20–33, doi: [10.1016/j.micron.2014.04.001](https://doi.org/10.1016/j.micron.2014.04.001), indexed in Pubmed: [25041828](https://pubmed.ncbi.nlm.nih.gov/25041828/).
88. Savkare SS, Narote SP. Automated system for malaria parasite identification. *IEEE.* 2015: 1–4, doi: [10.1109/iccict.2015.7045660](https://doi.org/10.1109/iccict.2015.7045660).
89. Savkare SS, Narote SP. Automatic detection of malaria parasites for estimating parasitemia. *Int J Comput Sci Secur IJCSS.* 2011; 5: 310.
90. Savkare SS, Narote SP. Automatic System for Classification of Erythrocytes Infected with Malaria and Identification of Parasite's Life Stage. *Procedia Technology.* 2012; 6: 405–410, doi: [10.1016/j.protcy.2012.10.048](https://doi.org/10.1016/j.protcy.2012.10.048).
91. Schroder S, Wittmann D, Drescher W, Roth V, Steinhage V, Cremers AB: The new key to bees: automated identification by image analysis of wings. *Pollinating Bees– Conserv Link Agric Nat Minist Environ Bras* 2002. https://www.researchgate.net/profile/Volker_Roth/publication/241318944_THE_NEW_KEY_TO_BEES_AUTOMATED_IDENTIFICATION_BY_IMAGE_ANALYSIS_OF_WINGS/links/0046352a5957f5816d000000.pdf (cited 2016 Oct 2).
92. Shih TK, Huang J-Y, Wang C-S, et al. An intelligent content-based image retrieval system based on color, shape and spatial relations; in: *Proceedings-National Science Council Republic Of China Part A Physical Science And Engineering.* 2001 [cited 2016 Oct 2], pp. 232–243.
93. Singh HK, Tomar SK, Maurya PK. Thresholding Techniques applied for Segmentation of RGB and multi-spectral images. *Proc Publ Int J Comput Appl.* 2012: 975–8887.
94. Singhai N, Shandilya S. A Survey On: “Content Based Image Retrieval Systems”. *Int J Comput Appl.* 2010; 4(2): 22–26, doi: [10.5120/802-1139](https://doi.org/10.5120/802-1139).
95. Soberón J, Peterson AT. Biodiversity informatics: managing and applying primary biodiversity data. *Philos Trans R Soc Lond B Biol Sci.* 2004; 359(1444): 689–698, doi: [10.1098/rstb.2003.1439](https://doi.org/10.1098/rstb.2003.1439), indexed in Pubmed: [15253354](https://pubmed.ncbi.nlm.nih.gov/15253354/).
96. Thiel SU, Wiltshire RJ, Davies LJ. Automated object recognition of blue-green algae for measuring water quality: a preliminary study. *Oceanogr Lit Rev.* 1996; 1: 85.
97. Trattner S, Greenspan H, Tepper G, et al. Automatic identification of bacterial types using statistical imaging methods. *IEEE Trans Med Imaging.* 2004; 23(7): 807–820, doi: [10.1109/TMI.2004.827481](https://doi.org/10.1109/TMI.2004.827481), indexed in Pubmed: [15250633](https://pubmed.ncbi.nlm.nih.gov/15250633/).
98. Underwood J, Dahlberg A, FitzPatrick S, et al. A STILE Project case study: The evaluation of a computer-based visual key for fossil identification. *Assoc Learn Technol J.* 1996; 4(2): 40–47, doi: [10.1080/0968776960040205](https://doi.org/10.1080/0968776960040205).
99. Ververidis D, Kotropoulos C. Fast and accurate sequential floating forward feature selection with the Bayes classifier applied to speech emotion recognition. *Signal Processing.* 2008; 88(12): 2956–2970, doi: [10.1016/j.sigpro.2008.07.001](https://doi.org/10.1016/j.sigpro.2008.07.001).
100. Vogt A, Cholewinski A, Shen X, et al. Automated image-based phenotypic analysis in zebrafish embryos. *Dev Dyn.* 2009; 238(3): 656–663, doi: [10.1002/dvdy.21892](https://doi.org/10.1002/dvdy.21892), indexed in Pubmed: [19235725](https://pubmed.ncbi.nlm.nih.gov/19235725/).
101. Wang J, Lin C, Ji L, et al. A new automatic identification system of insect images at the order level. *Knowledge-Based Systems.* 2012; 33: 102–110, doi: [10.1016/j.knosys.2012.03.014](https://doi.org/10.1016/j.knosys.2012.03.014).
102. Weeks P, O'Neill MA, Gaston KJ, et al. Species-identification of wasps using principal component associative memories. *Image Vis Comp.* 1999; 17(12): 861–866, doi: [10.1016/s0262-8856\(98\)00161-9](https://doi.org/10.1016/s0262-8856(98)00161-9).

103. Wen C, Guyer D, Li W. Local feature-based identification and classification for orchard insects. *Biosystems Eng.* 2009; 104(3): 299–307, doi: [10.1016/j.biosystemseng.2009.07.002](https://doi.org/10.1016/j.biosystemseng.2009.07.002).
104. Witten IH, Frank E: *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005. https://books.google.com/books?hl=en&lr=&id=QTnOcZJzIUoC&oi=fnd&pg=PR17&dq=Data+Mining:+Practical+machine+learning+tools+and+techniques&ots=3hnBcpZlQe&sig=jDee4hf43_u_bRcYZ7yoNzZtuZQ (cited 2016 Oct 2).
105. Yang M, Kpalma K, Ronsin J. A Survey of Shape Feature Extraction Techniques. *Pattern Recognition Techniques, Technology and Applications*. 2008: 43–90, doi: [10.5772/6237](https://doi.org/10.5772/6237).
106. Yang YS, Park DK, Kim HC, et al. Automatic identification of human helminth eggs on microscopic fecal specimens using digital image processing and an artificial neural network. *IEEE Trans Biomed Eng.* 2001; 48(6): 718–730, indexed in Pubmed: [11396601](https://pubmed.ncbi.nlm.nih.gov/11396601/).
107. Yu L, Liu H: Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution; in : *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*. 2003, [cited 2016 Oct 2], pp 856–863.
108. Yu X, Wang J, Kays R, et al. Automated identification of animal species in camera trap images. *EURASIP J Image Video Process.* 2013; 2013(1): 52, doi: [10.1186/1687-5281-2013-52](https://doi.org/10.1186/1687-5281-2013-52).
109. Yuan G, Hasler N, Klette R, Rosenhahn B: *Understanding Tracks of Different Species of Rats* (2006). CITR, The University of Auckland, New Zealand, 2006. <https://researchspace.auckland.ac.nz/handle/2292/2783> (cited 2016 Oct 2).
110. Yue J, Li Z, Liu Lu, et al. Content-based image retrieval using color and texture fused features. *Math Comput Model.* 2011; 54(3-4): 1121–1127, doi: [10.1016/j.mcm.2010.11.044](https://doi.org/10.1016/j.mcm.2010.11.044).
111. Zhan M, Crane MM, Entchev EV, et al. Automated Processing of Imaging Data through Multi-tiered Classification of Biological Structures Illustrated Using *Caenorhabditis elegans*. *PLoS Comput Biol.* 2015; 11(4): e1004194, doi: [10.1371/journal.pcbi.1004194](https://doi.org/10.1371/journal.pcbi.1004194), indexed in Pubmed: [25910032](https://pubmed.ncbi.nlm.nih.gov/25910032/).
112. Zhang D, Lu G. Review of shape representation and description techniques. *Pattern Recognition.* 2004; 37(1): 1–19, doi: [10.1016/j.patcog.2003.07.008](https://doi.org/10.1016/j.patcog.2003.07.008).