

A Study of Image Analysis Algorithms for Segmentation, Feature Extraction and Classification of Cells

Margarita Gamarra^{1,2,*}, Eduardo Zurek², Homero San-Juan³

¹ Dept. Electronic Eng., Universidad Autónoma del Caribe, COLOMBIA

² Dept. System Eng., Universidad del Norte, Barranquilla, COLOMBIA

³ Dept. Medicine, Universidad del Norte, Barranquilla, COLOMBIA

*Corresponding Author: mrgamarra22@gmail.com

Citation: Gamarra, M., Zurek, E. and San-Juan, H. (2017). A Study of Image Analysis Algorithms for Segmentation, Feature Extraction and Classification of Cells. *Journal of Information Systems Engineering & Management*, 2(4), 20. <https://doi.org/10.20897/jisem.201720>

Published: August 30, 2017

ABSTRACT

Recent advances in microscopy and improvements in image processing algorithms have allowed the development of computer-assisted analytical approaches in cell identification. Several applications could be mentioned in this field: Cellular phenotype identification, disease detection and treatment, identifying virus entry in cells and virus classification; these applications could help to complement the opinion of medical experts. Although many surveys have been presented in medical image analysis, they focus mainly in tissues and organs and none of the surveys about image cells consider an analysis following the stages in the typical image processing: Segmentation, feature extraction and classification. The goal of this study is to provide comprehensive and critical analyses about the trends in each stage of cell image processing. In this paper, we present a literature survey about cell identification using different image processing techniques.

Keywords: microscope technology, segmentation, feature extraction, cell image, image processing

INTRODUCTION

Nowadays several microscopy modalities have been developed and even if we limit ourselves to the visible portion of the electromagnetic spectrum we find numerous techniques such as fluorescence microscopy, differential interference contrast microscopy, phase contrast microscopy, dark field microscopy, confocal microscopy and others (Göröcs and Ozcan 2013).

Millions of cells are present in thousands of images obtained by different microscopy technologies and in the presence of millions of cells the human visual classification becomes infeasible. Although these technologies in microscopy offer high-resolution images, the big data and pattern not recognized by human experts are the main factors to implement image processing and image analysis techniques in cell identification. Machine learning and data mining have the potential to objectively and effectively analyze the massive amounts of image data (Shamir et al. 2010).

In addition, qualitative and quantitative characterization of cell images is important for clinical applications (e.g., vaccine development and diagnosis and treatment of disease), biological research (e.g., to understand the input mechanisms of the virus in cells) and application of improved techniques in digital image processing (Ketteler and Kriston-Vizi 2016).

In this review, we have done a literature search in different databases, including medical, biological and engineering databases. Detailed searches of ISI Web of Science, Scopus, ScienceDirect, IEEE Xplore, PubMed, BioMed, Medline and Google Scholar, were conducted using cell images processing, technologies of microscopy

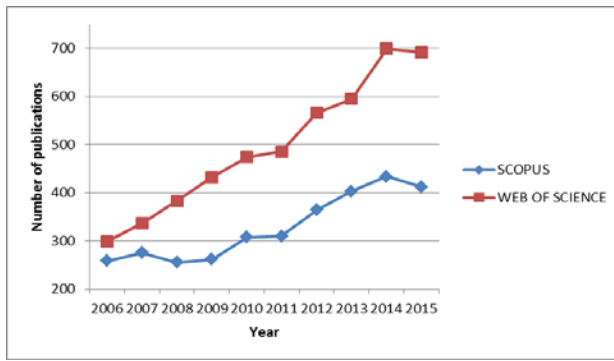


Figure 1. Evolution of the number of publications per year retrieved from the keywords “cell image identification” in Web of Science and Scopus, the last ten years

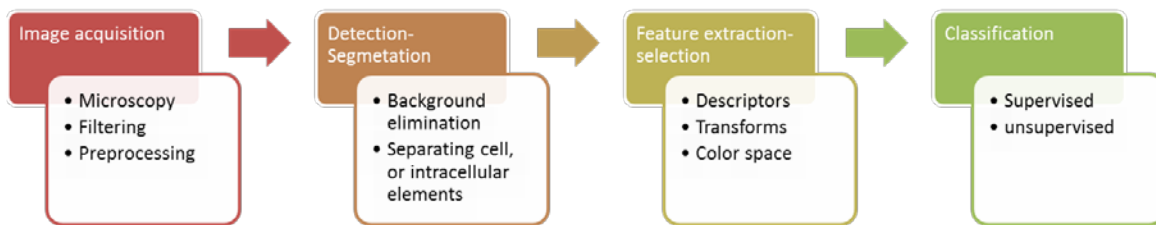


Figure 2. Scheme of the image analysis used for cell identification

and cell image classifier. The search included papers published in English, using a temporal window of ten years (since 2006 to 2016). Titles and abstracts of potentially relevant studies were selected. Afterwards, we ordered the full texts of these studies and evaluated whether they would be included. Initially, we identify about 200 records through database searching, but after remove duplicated registers and select the ones closest to the topic, we included only 70 related works. We have organized this paper to follow the image analysis procedures for cell identification and generally for all imaging processing. Then, only a few set of representative papers are described in more detail in each stage of image processing.

The application of these image analysis techniques is confirmed by the growing trend in the number of publications shown in [Figure 1](#). The figure reproduces the evolution of the number of references retrieved by the application of the specific keywords “cell image identification” since 2006 in the academic search sites Scopus and Web of Science, which collects medicine, engineering and technological publications. This behavior can be attributed to the increased use of the new technologies in microscopy supported by rapid growth in software development, management of big data and enhancement of imaging techniques.

The aim of this review is to provide a general vision of the implementation and progress of image analysis applied to cell identification suitable for biological and medical studies. Previously the authors published a literature survey about cell identification in ([Gamarra, Zurek, and San Juan 2017](#)).

We have organized this paper to follow the image analysis procedures for cell identification and generally for all imaging processing. Some researches in virus identification have been included, because the techniques of image processing are very similar to cell identification, although the image acquisition is performed in a different scale. The scheme in [Figure 2](#) represents the main stages of the image analysis and certain characteristics used in cell identification process in each stage.

SEGMENTATION

The images obtained from a microscope show not only a unit of cell or virus but hundreds of cells are contained by each image. This makes the first operation in many cases, is the segmentation of elements in the image. The main motivation for detecting and segmenting the observed structures is the need for counting of objects, generally cells or cell nuclei ([Kong et al. 2011](#)), to find the rate of infected cells by virus, evaluate the growth rate of microorganisms, calculation of blood cells ([Seigel et al. 2012](#)), etc. Cell counts can have diagnostic significance for some cancerous conditions ([Lim, Mashor, and Hassan 2012](#)) or virus infection.

Several techniques for segmentation are well known in image processing, but not all are useful in the microscopic range. In most biological images, cells touch each other, causing the simple, fast algorithms used in other image processing cases to fail ([Carpenter et al. 2006](#); [Zhang, Wang, and Shi 2009](#)). The segmentation stage has applications itself like quantification of cell and cell infected by virus, but it can also be used for a later processing with identification and characterization objectives. Depending on the element to segment (i.e.

Table 1. Summary of segmentation techniques used in cell image analysis

Technique	Description	Weaknesses	References
Thresholding	It is implemented in color or gray scale images. This technique is based on the histogram. Used as a complementary process with other methods. Useful in cell nuclei segmentation process.	Difficulty to find an adequate thresholding. This technique requires the foreground and background have different intensity values.	(Matula et al. 2009), (Liao et al. 2015), (Dogantekin, Avci, and Erkus 2013), (Stoklasa, Majtner, and Svoboda 2014), (Filipczuk, Krawczyk, and Woźniak 2013)
Region-based segmentation	These techniques operate iteratively by grouping together pixels which have similar values. The watershed transform is a region-based segmentation technique. It produces a division of the image in separated regions.	It could produce unsmooth boundary for the extracted object.	(Lim, Mashor, and Hassan 2012), (Karvelis et al. 2006), (Huang 2010), (Tonti et al. 2015), (Huang and Murphy 2004), (Mouelhi et al. 2013)
Edge-based segmentation	An edge filter is applied to the image and pixels are classified as edge or non-edge. These are usually detected by the first or second order derivatives method.	False edges could be included, and then post-processing operations are required.	(Lim, Mashor, and Hassan 2012), (Liao et al. 2015), (Gopinath and Bovik 2012), (Dogantekin, Avci, and Erkus 2013), (Mao et al. 2014), (Castañón et al. 2007), (Leonard et al. 2015)
Energy-based segmentation	These techniques aim to minimize an energy function when the image is segmented correctly. It includes algorithms like graph-cut (Yi and Moon 2012), live wire, active contour and level sets.	It depends on the choice of the initial contour which has to be close to the desired minima.	(Zhang, Wang, and Shi 2009), (Bergeest and Rohr 2012), (Tarnawski et al. 2013), (Liao et al. 2015), (Mouelhi et al. 2013)
Clustering	These techniques are used in the first exploratory data analysis and to group patterns that are similar. Sometimes they are combined with other techniques.	It can require a post-supervision.	(Kong et al. 2011), (Tonti et al. 2015), (Abeysekera et al. 2014), (Filipczuk, Krawczyk, and Woźniak 2013)

separating cell and the background, nucleus of membrane or intracellular elements) different techniques are implemented such as thresholding and fuzzy c-means clustering. However, thresholding tends to have good results only on uniform images and does not produce consistent results if there is variability within image sets. Watershed algorithms have the same problem (Lim, Mashor, and Hassan 2012) (Karvelis et al. 2006). The morphological techniques work only when the concavity degree of the overlapping cells is large enough (Huang 2010), which is not usually the case. Thus, other algorithms have been proposed to improve the accuracy of the segmentation process, including improved watershed algorithms (Tonti et al. 2015). In **Table 1** we present the most common segmentation techniques applied to cell images, a brief description and the weaknesses existing for the application in cell image analysis.

In (Matula et al. 2009) authors have developed an image analysis approach that comprises (i) a gradient-based thresholding scheme for cell nuclei segmentation which does not require subsequent postprocessing steps for separation of clustered nuclei, (ii) quantification of the virus signal in the neighborhood of cell nuclei, (iii) localization of regions with transfected cells by combining model-based circle fitting and grid fitting, (iv) cell classification as infected or noninfected, and (v) image quality control (e.g., identification of out-of-focus images). Due to it is difficult to find a single global threshold suitable for all nuclei, authors propose an edge-based approach which analyzes gradient magnitude images instead of thresholding the original image intensities. To distinguish between interior and exterior cell nuclei parts and to improve the segmentation result for clustered cells, authors calculate second derivatives of an image and locate zero-crossings of the Laplacian operator. The main idea is to detect pixels on the bright side of edges that have a large gradient magnitude and to subsequently morphologically process the result to obtain the final segmentation. With this approach, authors obtained a 97% in accuracy for the segmentation process.

The approach proposed in (Bergeest and Rohr 2012) is based on active contours and level sets for cell nuclei segmentation in fluorescence microscopy images. Authors have developed two approaches. (i) The three-step approach employs two convex energy functionals, which are convex formulations of the region based Chan–Vese (Chan and Vese 2001) functional and the Bayesian functional. By applying the convex functionals subsequently in three steps is possible deal with global intensity inhomogeneities and neighboring cells. The two-step approach employs the convex Bayesian functional and the convex region-scalable fitting energy functional, and requires only two subsequent steps. In the first step, authors use the region-scalable fitting energy functional, which can cope

with spatially local mean intensity values in comparison to the region-based Chan–Vese functional and the Bayesian functional which are based on global mean intensity values. In the second step, authors use the Bayesian functional which allows dealing with inhomogeneities at the border of cell nuclei. From the experiments it turned out that both the three-step approach and the two-step approach can cope with images of different cell types. Authors also found that the two-step approach yields superior results compared to the three-step approach and compared to previous approaches.

Authors in (Tarnawski et al. 2013) propose a new algorithm for segmenting and tracking of clustered cells. The algorithm assumes that the objects to be tracked or segmented are of elliptical shape. Based on H -minima transform, the key is to find the optimal b (given depth. Typically, it is between 2 and 10). Authors propose instead of finding a global b for the entire image to determine it for each object separately. Authors assume that nuclei have elliptical shapes, and they use ellipse fitting for finding the best choice of b for each object. Concretely, they apply H -minima transform with a fixed b and the watershed algorithm. The results show that the proposed algorithm for segmentation and tracking significantly outperforms *CellProfiler* and *MTrack2* software. It slightly outperforms *LSetCellTracker* (the latter one is only better under specificity measure), which moreover is much more sensitive to changes in parameters. The measured values for sensitivity, specificity and precision were higher than 93%, 74% and 89%, respectively.

The proposed method in (Liao et al. 2015) mainly consists of preprocessing, polygon approximation, bottle neck detection, ellipse fitting, cell segmentation, and edge modification. Firstly, a binary image of cell contours is obtained by preprocessing, followed by a polygon approximation to extract feature points of cell contours. Secondly, for each connected cell region, bottleneck detection and ellipse fitting are used to judge whether it is a single cell or overlapping cells, and further to detect the splitting points for overlapping cells. Thirdly, according to the splitting points, one cell is separated from the overlapping cells by a fitted ellipse. Finally, the remaining edge is patched up to form a new closed contour by edge modification. The proposed method separates cells from overlapping cells iteratively one by one, and the accuracies for blood and fluorescence cell segmentations are 92% and 90%, respectively.

FEATURE EXTRACTION

After a segmenting process and preprocessing operations to adequate the images is necessary extract features of the cells, virus or other microorganism. This stage allows to characterize the cells or virus itself or to perform a posterior classification. Depending of the elements to identify several types of features are extracted: morphological (Gopinath and Bovik 2012), texture, invariant moments, co-occurrence matrix (Liu and Liu 2013), geometric and appearance property (wavelet features (Liu et al. 2013), Zernike moments features (Apostolopoulos, Tsinopoulos, and Dermatas 2013), Haralick features, region property features, shape descriptor features) (Wang et al. 2006), entropy values (Dogantekin, Avci, and Erkus 2013), higher order spectral (Ong and Chandran 2005) and color (Jaiswal et al. 2003; Huang et al. 2006; Karkanis et al. 2003).

A summary of features used in cells identification is listed in **Table 2**. These features were compiled from a comprehensive literature search on cells image analysis. In addition, various statistics measures are also commonly calculated for all vector features: the mean, median, minimum, maximum, standard deviation, texture histogram

Table 2. Summary of features used in cell image analysis

Category	Features	References
Size and structure	Area	(Kuznetsov and McPherson 2011), (Chen et al. 2013), (Kuznetsov et al. 2012), (Gopinath and Bovik 2012), (Bocklitz et al. 2014), (Ashcroft et al. 2011), (Abbas, Dijkstra, and Heskes 2014)
	Elliptical features: major and minor axis length, eccentricity, orientation, elliptical deviation.	
	Convex area.	
	Boundary features: perimeter, radio, perimeter curvature.	
Texture	Other shape features: equivalent diameter, sphericity, inertia shape.	(Liu and Liu 2013), (Liu et al. 2013), (Wang et al. 2006), (Dogantekin, Avci, and Erkus 2013), (Ong and Chandran 2005), (Gertych et al. 2015), (Abeysekera et al. 2014), (Mao et al. 2014), (dos Santos et al. 2015), (Kayaalti et al. 2014), (Stoklasa, Majtner, and Svoboda 2014)
	Co-occurrence matrix features: inertia, energy, entropy, homogeneity, maximum probability.	
	Wavelet features.	
Color	Fractal dimension.	(Jaiswal et al. 2003), (Huang et al. 2006), (Karkanis et al. 2003), (Filipczuk, Krawczyk, and Woźniak 2013)
	Color coherence vectors (CCV), border/Interior classification (BIC), color correlograms	

(Gertych et al. 2015), skewness, and kurtosis. Other authors proposed a combination, modification or new variants of several well-known features.

Authors in (Bocklitz et al. 2014) use the image moment method in combination with atomic force microscopy (AFM) to discriminate viruses on a single-particle level. The images recorded by AFM can be analyzed using image analysis approaches, as the measured quantity can be converted to a gray scale. The method best suited for such isolated virion images is the image moment method (Flusser, Suk, and Zitová 2009). In this research an automatic segmentation is not used, instead the region of interest (sub-scan) is chosen manually then the analysis presented is not corrupted by possible artifacts produced by an automatic segmentation. After the pre-treatment was finished, features from the virion sub-scan were extracted, which included Pseudo-Zernike-Moments (PZ-moments), area, volume and the height deciles. As the values differ strongly in their order of magnitude a principal-component-analysis (PCA) with centering and scaling to unit variance of every feature was applied. Subsequently, a dimension reduction to 50 PCs was carried out and a linear-discriminant-analysis (LDA) was applied to build up a classification model for the virus species. The accuracy of this classification model was 97%.

The research presented in (Abeysekera et al. 2014) focuses on statistical feature extraction, which is an approach that utilizes statistical operations such as zoning, projection, profiling, histogram and distance moments (Kononenko and Kukar 2007). The aim of this research is to develop a computational pathology framework for detecting spongiosis in a stained histopathological specimen using Multi-Spectral Imaging (MSI) and customized feature extraction methods. The proposed set of features is used to train a support vector machines (SVM) and sparse linear discriminant analysis (SLDA) classifiers to detect the areas of spongiosis. The results show that SLDA outperforms SVM quite significantly.

Authors in (Mao et al. 2014) present other feature extraction technique applied in virus plaque. Viral plaque is a visible structure formed by phage infection and propagation within a cell. Particularly, the parameters such as the number, diameter, and area of plaques are needed to accurately measure for further comparison or classification of plaques. In this paper an effective and novel method for the detection and segmentation of plaques is presented by means of histogram of oriented gradient (HOG) features (Dalal and Triggs 2014) and support vector machine (SVM). The HOG was first used to calculate the occurrence of gradient direction in the local patch of an image. It captured edge and gradient structures, which are characteristic of local shapes, while it performed photometric and geometric transformations. Extensive experimentation showed that the method was feasible in detecting and segmenting plaques and the performance was satisfied by comparing it to the ones measured by experts manually.

In the paper (dos Santos et al. 2015) authors present a new ensemble of descriptors for the classification of transmission electron microscopy images of viruses that is based on texture analysis. A set of six well-assessed texture descriptors, namely Local Binary Pattern (LBP) (Ojala, Pietikainen, and Maenpaa 2002), Local Ternary Pattern (LTP) (Tan and Triggs 2010), Dense LBP (DLBP) (Ylioinas et al. 2013), Rotation Invariant co-occurrence among LBP (RI) (Nosaka, Yasuhiro, and Kasuhiro 2012), Local Phase Quantization (LPQ) (Ojansivu and Heikkilä 2008), and LBP Histogram Fourier (LHF) (Ahonen et al. 2009), are combined with innovative approaches to improve their performances in virus classification. To generate new variants of the aforementioned descriptors, different approaches are applied: (i) the Edge approach (ED) (Abdesselam 2013), which extracts the textural information from specific regions of the image instead of from the original image; (ii) the Bag of Features (BOF) scheme (Nowak, Jurie, and Triggs 2006), which is used to build vocabularies of the most representative patterns; and (iii) Multi-Quinary coding (MQ) (Paci et al. 2013). In all experiments, SVM (using both linear and radial basis function kernels) is the base classifier. We observe that ED improves the performance of the single standard descriptors, while BOF is particularly effective in fusion by sum rule with the standard application of texture descriptors (i.e., when they are extracted from the entire image). In contrast, the combination of MQ with LHF, DLBP, and RI does not improve virus classification. Using these results, authors suggest a new ensemble of descriptors called NewF, whose accuracy in virus classification is 85.7%, outperforming previous methods proposed in the literature for the same task.

In many cases the features extracted develop a large feature vector, including irrelevant features. This lot of features could complicate the classification process. Then it is necessary to include a feature selection stage before classification.

Some well-known and commonly used methods of linear dimensionality reduction are Principal Component Analysis (PCA), Independent Component Analysis (ICA), and Linear Discriminant Analysis (LDA). On the other hand, several heuristic algorithms have been developed, which use the hit rate classification as the optimality criterion. These methods include the sequential search methods: sequential forward selection (SFS) and sequential backward selection (SBS) (Pudil, Novovičová, and Kittler 1994; Castañón et al. 2007). Other optimization techniques used in features selection are genetic algorithms. A taxonomy of feature selection algorithms is presented in (Jain and Zongker 1997).

Table 3. Summary of classifier used in cell image analysis

Category	Classifier	References
Statistical and classical	Probabilistic: Bayesian classifier, probabilistic linear discriminant analysis.	(Dogantekin, Avci, and Erkus 2013), (Karkanis et al. 2003), (Gertych et al. 2015), (Ashcroft et al. 2011), (Stoklasa, Majtner, and Svoboda 2014), (Nanni and Lumini 2007), (Filipcuk, Krawczyk, and Woźniak 2013)
	Non-probabilistic: support vector machine, K-nearest neighbor and linear discriminant analysis.	
Artificial intelligence	Artificial Neural Networks, Fuzzy logic, Genetic Algorithms, decision tree.	(Filipcuk, Krawczyk, and Woźniak 2013), (Ashcroft et al. 2011), (dos Santos et al. 2015), (Kayaalti et al. 2014), (Glezakos et al. 2010), (Keltch, Lin, and Bayrak 2014), (Sagonas et al. 2013)
Boosting	Joint boosting, gentle boosting, adaptive boosting.	(Abbas, Dijkstra, and Heskes 2014), (Li et al. 2012), (Huang and Murphy 2004)

CLASSIFICATION

For cells image analysis, one of the primary considerations in the choice of a classifier is its ability to deal with highly dense datasets and additionally a large feature vector. According with (Jones et al. 2009), there are three steps involved in classification. The first step is segmentation and feature extraction. The second step concerns the training of classification models on a training set and their performance evaluation with cross-validation. The training set is a subset of a few thousand cells visually classified by a biologist. The third step develops the classification of whole screen using the best performing classifier from step 2.

Some well know classifiers are used for this image analysis. A summary of classifier used in cell identification is listed in **Table 3**. Additionally, the classifier can be divided into two basic groups: supervised and unsupervised (Sonka, Hlavac, and Boyle 2007). In the supervised learning the samples labeled of each class are provided; in unsupervised learning the samples are not labeled and the classifier can try to cluster the data into different groups.

Some researches use intelligent algorithms to classify the cell objective (dos Santos et al. 2015; Glezakos et al. 2010; Keltch, Lin, and Bayrak 2014), soft computing (Kayaalti et al. 2014; Çalışır and Dogantekin 2011), k-means (Ashcroft et al. 2011), decision tree (Leonard et al. 2015) or a combination of two or more techniques (Li et al. 2012; Huang and Murphy 2004). One case is presented in (Dogantekin, Avci, and Erkus 2013), where authors proposed in the classification stage an Adaptive Network Based on Fuzzy Inference System (ANFIS) to automatically classify RNA viruses. In the feature extraction stage, each of the RNA virus images is rotated 15° , between 0° and 165° . These rotated RNA virus images are then scaled to 10 different sizes. Moreover, three different entropy values, i.e., the norm, logarithmic energy and entropy threshold values are calculated for each of these RNA virus images. The feature vector is given as input to the ANFIS classifier, Artificial Neural Networks (ANN) classifier and Fuzzy c-means (FCM) cluster. Finally, the test stage is performed to evaluate the correct classification ratio of the Entropy-ANFIS algorithm for the RNA virus images. The correct classification ratio was determined as 95% using the proposed Entropy-ANFIS method.

In (Abbas, Dijkstra, and Heskes 2014) a comparative study of cell classifier is presented. Authors present a comparative study of computational performance (accuracy and cross-validation time) of gentle boosting, joint boosting CellProfiler Analyst (CPA), support vector machines (linear and radial basis function) and linear discriminant analysis (LDA) on two data sets of HT29 and HeLa cancer cells. Authors also explore how performance and computational time vary with a different number of phenotypes. The extracted features consist of geometric (extension, eccentricity, axis lengths, size and size ratio between cell and nucleus), Haralick (angular moments, contrast, correlation, variance and entropy) and Zernike features. Authors use 20-fold cross-validation with stratified sampling on the class variables. The results show that the performance of classification methods increases with a decrease in the number of phenotypes for both data sets and the cross-validation time increases with the number of phenotypes. This study finds that the difference in performance is small between SVM (linear) and SVM (RBF) but that SVM (linear) is faster than SVM (RBF) on both data sets. Then authors propose SVM (linear) for iterative improvement of the training data and SVM (RBF) for the final classifier to classify all unlabeled cells in the whole data set.

In the research presented in (Stoklasa, Majtner, and Svoboda 2014) authors use k-nearest neighbor (k-NN) to classify Human Epithelial (HEp2) cells. The classifier is able to categorize pre-segmented images of HEp-2 cells into 6 classes. The core of this engine consists of the following image descriptors: Haralick features, Local Binary Patterns, Scale Invariant Feature Transform (SIFT), surface description and a granulometry-based descriptor. The proposed classifier is k-NN based and was implemented using MESSIF framework (Batko et al. 2008). In comparison with SVM, the advantages of k-NN based approach in the implicit support of multi-class division and

the ability to natively combine different weighted feature vectors. Presented results show that the performance is highly influenced by the intra-and inter-image similarity. When inter-image differences are suppressed by using leave-one-out-cross-validation on cell- Level using the whole dataset, classifier achieves more than 93% accuracy.

In some cases due to multiple image scales at which relevant information may be extracted from cell images, some authors have proposed the use of multiple classifier system as opposed to a single classifier (Gurcan et al. 2009; Nanni and Lumini 2007; Lino et al. 2016). These systems perform information fusion of classification decisions at different levels overcoming limitations of traditional approaches based on single classifiers (Woźniak, Graña, and Corchado 2014). Likewise a combination of several algorithms has been used in medical image fusion: process of registering and combining multiple images from single or multiple imaging modalities to improve the imaging quality and reduce randomness and redundancy in order to increase the clinical applicability of medical images for diagnosis and assessment of medical problems (James and Dasarathy 2014).

An example of classifier ensemble is presented in (Filipczuk, Krawczyk, and Woźniak 2013) applied in cytological images. Authors propose a segmentation method based on the combination of adaptive thresholding in grayscale and pixel classification in color space using one of four clustering algorithms: K-means, fuzzy C-means, Gaussian mixture model (Dempster, Laird, and Rubin 1976), and competitive neural networks. Competitive learning neural networks, although computationally costly, returned a best results and features extracted on its basis contributed to the lowest classification error. Classification is carried out with the usage of a classifier ensemble based on the Random Subspaces approach (Ho 1998). To boost its effectiveness, authors use a linear combination of the support functions returned by the individual classifiers in the ensemble. The ensemble must consist of classifiers that are able to output support functions as Naive Bayes, Neural Networks, or Support Vector Machines. The implementation of such a fuser was motivated by the fact that not all of the base classifiers created by the Random Subspace method are of the same quality – while it is a good method for assuring diversity it cannot always ensure the high individual accuracy. By using a trained fuser, authors controlled the influence that each base classifier had on the final decision. This resulted in a high-quality ensemble, which outperformed other committees, commonly used in the medical decision support task.

CONCLUSIONS

Cell image analysis techniques in the last years have been presented in this review. The development of new technologies in microscopy has increased the study of cell and virus using the image analysis. Well known techniques in image processing are used to identify phenotypes, infected cells, new virus and in general a quantification and characterization of cell images. Although the image analysis is finally carried out by human experts, it is necessary to use automatic processing due to the big data that the images contain.

Each stage of the image processing has a wide field of research: segmentation and preprocessing, feature extraction and feature selection, classification and improving of image quality. In these fields are evaluated two important indicators: accuracy and execution time. Both indicators are critical because of the amount of information that must be processed.

Although the advance in cell image analysis has had a high increase the last years, there exist technical challenges in image acquisition resulting from image noise, high cost of imaging and increased computational complexity with increasing image amount. Facing this, new techniques or combination of existing are implemented aim to improve the performance in this image analysis.

REFERENCES

- Abbas, S.S., Dijkstra, T.M.H. and Heskes, T. (2014). A Comparative Study of Cell Classifiers for Image-Based High-Throughput Screening. *BMC bioinformatics*, 15(1), p. 342. <https://doi.org/10.1186/1471-2105-15-342>
- Abdesselam, A. (2013). Improving Local Binary Patterns Techniques by Using Edge Information. *Lecture Notes on Software Engineering*, 1(4), pp. 360–363. <https://doi.org/10.7763/LNSE.2013.V1.77>
- Abeysekera, S., Ooi, M.P.-L., Kuang, Y.C., Tan, C.P. and Hassan, S.S. (2014). Detecting Spongiosis in Stained Histopathological Specimen Using Multispectral Imaging and Machine Learning. In *2014 IEEE Sensors Applications Symposium (SAS)*. IEEE, pp. 195–200. <https://doi.org/10.1109/SAS.2014.6798945>
- Ahonen, T., Matas, J., He, C. and Pietikäinen, M. (2009). Rotation Invariant Image Description with Local Binary Pattern Histogram Fourier Features. In A.-B. Salberg, J. Y. Hardeberg, and R. Jenssen, eds. *Image Analysis. Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer Berlin Heidelberg, p. Image Analysis. <https://doi.org/10.1007/978-3-642-02230-2>

- Apostolopoulos, G., Tsinopoulos, S.V. and Dermatas, E. (2013). A Methodology for Estimating the Shape of Biconcave Red Blood Cells Using Multicolor Scattering Images. *Biomedical Signal Processing and Control*, 8(3), pp. 263–272. <https://doi.org/10.1016/j.bspc.2012.11.002>
- Ashcroft, F.J., Newberg, J.Y., Jones, E.D., Mikic, I. and Mancini, M.A. (2011). High Content Imaging-Based Assay to Classify Estrogen Receptor- α Ligands Based on Defined Mechanistic Outcomes. *Gene*, 477(1–2), pp. 42–52. <https://doi.org/10.1016/j.gene.2011.01.009>
- Batko, M., Dohnal, V., Novák, D. and Zezula, P. (2008). Metric Similarity Search Implementation Framework (MESSIF). [online]. Available from: <https://www.muni.cz/vyzkum/publikace/836626>
- Bergeest, J.-P. and Rohr, K. (2012). Efficient Globally Optimal Segmentation of Cells in Fluorescence Microscopy Images Using Level Sets and Convex Energy Functionals. *Medical Image Analysis*, 16(7), pp. 1436–1444. <https://doi.org/10.1016/j.media.2012.05.012>
- Bocklitz, T., Kämmer, E., Stöckel, S., ... Popp, J. (2014). Single Virus Detection by Means of Atomic Force Microscopy in Combination with Advanced Image Analysis. *Journal of structural biology*, 188(1), pp. 30–38. <https://doi.org/10.1016/j.jsb.2014.08.008>
- Carpenter, A.E., Jones, T.R., Lamprecht, M.R., ... Sabatini, D.M. (2006). CellProfiler: Image Analysis Software for Identifying and Quantifying Cell Phenotypes. *Genome biology*, 7(10), p. R100. <https://doi.org/10.1186/gb-2006-7-10-r100>
- Castañón, C.A.B., Fraga, J.S., Fernandez, S., Gruber, A. and da F. Costa, L. (2007). Biological Shape Characterization for Automatic Image Recognition and Diagnosis of Protozoan Parasites of the Genus Eimeria. *Pattern Recognition*, 40(7), pp. 1899–1910. <https://doi.org/10.1016/j.patcog.2006.12.006>
- Chan, T.F. and Vese, L.A. (2001). Active Contours without Edges. *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*, 10(2), pp. 266–77. <https://doi.org/10.1109/83.902291>
- Chen, S.W., Odorico, M., Meillan, M., ... Pellequer, J.-L. (2013). Nanoscale Structural Features Determined by AFM for Single Virus Particles. *Nanoscale*, 5(22), pp. 10877–86. <https://doi.org/10.1039/c3nr02706f>
- Çalışır, D. and Dogantekin, E. (2011). A New Intelligent Hepatitis Diagnosis System: PCA–LSSVM. *Expert Systems with Applications*, 38(8), pp. 10705–10708. <https://doi.org/10.1016/j.eswa.2011.01.014>
- Dalal, N. and Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. IEEE, pp. 886–893. <https://doi.org/10.1109/CVPR.2005.177>
- Dempster, A.P., Laird, N.M. and Rubin, D.B. (1976). Maximum Likelihood from Incomplete Data via the EM Algorithm. <https://dash.harvard.edu/handle/1/3426318>
- Dogantekin, E., Avci, E. and Erkus, O. (2013). Automatic RNA Virus Classification Using the Entropy-ANFIS Method. *Digital Signal Processing*, 23(4), pp. 1209–1215. <https://doi.org/10.1016/j.dsp.2013.01.011>
- Filipczyk, P., Krawczyk, B. and Woźniak, M. (2013). Classifier Ensemble for an Effective Cytological Image Analysis. *Pattern Recognition Letters*, 34(14), pp. 1748–1757. <https://doi.org/10.1016/j.patrec.2013.04.030>
- Flusser, J., Suk, T. and Zitová, B. (2009). *Moments and Moment Invariants in Pattern Recognition*. Moments and Moment Invariants in Pattern Recognition Jan Flusser Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Prague, Czech Republic Tomáš Suk Institute of Information Theory and Automation, Academy of Science. <https://doi.org/10.1002/9780470684757>
- Gamarra, M., Zurek, E. and San Juan, H. (2017). Cell Identification Using Image Analysis: A Literature Survey. In *Information Systems and Technologies (CISTI), 2017 12th Iberian Conference*. pp. 890–896.
- Gertych, A., Ing, N., Ma, Z., ... Knudsen, B.S. (2015). Machine Learning Approaches to Analyze Histological Images of Tissues from Radical Prostatectomies. *Computerized Medical Imaging and Graphics: The Official Journal of the Computerized Medical Imaging Society*, 46 Pt 2, pp. 197–208. <https://doi.org/10.1016/j.compmedimag.2015.08.002>
- Glezakos, T.J., Moschopoulou, G., Tsiligiridis, T.A., Kintzios, S. and Yialouris, C.P. (2010). Plant Virus Identification Based on Neural Networks with Evolutionary Preprocessing. *Computers and Electronics in Agriculture*, 70(2), pp. 263–275. <https://doi.org/10.1016/j.compag.2009.09.007>
- Gopinath, A. and Bovik, A.C. (2012). Automatic Feature Extraction and Statistical Shape Model of the AIDS Virus Spike. *IEEE Transactions on Bio-Medical Engineering*, 59(12), pp. 3386–95. <https://doi.org/10.1109/TBME.2012.2215858>
- Göröcs, Z. and Ozcan, A. (2013). On-Chip Biomedical Imaging. *IEEE Reviews in Biomedical Engineering*, 6, pp. 29–46. <https://doi.org/10.1109/RBME.2012.2215847>
- Gurcan, M.N., Boucheron, L.E., Can, A., ... Yener, B. (2009). Histopathological Image Analysis: A Review. *IEEE Reviews in Biomedical Engineering*, 2, pp. 147–71. <https://doi.org/10.1109/RBME.2009.2034865>
- Ho, T.K. (1998). The Random Subspace Method for Constructing Decision Forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8), pp. 832–844. <https://doi.org/10.1109/34.709601>

- Huang, J. (2010). An Improved Algorithm of Overlapping Cell Division. In *2010 International Conference on Intelligent Computing and Integrated Systems*. IEEE, pp. 687–691. <https://doi.org/10.1109/ICISS.2010.5655507>
- Huang, K. and Murphy, R.F. (2004). Boosting Accuracy of Automated Classification of Fluorescence Microscope Images for Location Proteomics. *BMC Bioinformatics*, 5(1), p. 78. <https://doi.org/10.1186/1471-2105-5-78>
- Huang, X., El-Sayed, I.H., Qian, W. and El-Sayed, M.A. (2006). Cancer Cell Imaging and Photothermal Therapy in the near-Infrared Region by Using Gold Nanorods. *Journal of the American Chemical Society*, 128(6), pp. 2115–2020. <https://doi.org/10.1021/ja057254a>
- Jain, A. and Zongker, D. (1997). Feature Selection: Evaluation, Application, and Small Sample Performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2), pp. 153–158. <https://doi.org/10.1109/34.574797>
- Jaiswal, J.K., Mattoussi, H., Mauro, J.M. and Simon, S.M. (2003). Long-Term Multiple Color Imaging of Live Cells Using Quantum Dot Bioconjugates. *Nature Biotechnology*, 21(1), pp. 47–51. <https://doi.org/10.1038/nbt767>
- James, A.P. and Dasarathy, B. V. (2014). Medical Image Fusion: A Survey of the State of the Art. *Information Fusion*, 19, pp. 4–19. <https://doi.org/10.1016/j.inffus.2013.12.002>
- Jones, T.R., Carpenter, A.E., Lamprecht, M.R., ... Sabatini, D.M. (2009). Scoring Diverse Cellular Morphologies in Image-Based Screens with Iterative Feedback and Machine Learning. *Proceedings of the National Academy of Sciences of the United States of America*, 106(6), pp. 1826–31. <https://doi.org/10.1073/pnas.0808843106>
- Karkanis, S.A., Iakovidis, D.K., Maroulis, D.E., Karras, D.A. and Tzivras, M. (2003). Computer-Aided Tumor Detection in Endoscopic Video Using Color Wavelet Features. *IEEE Transactions on Information Technology in Biomedicine*, 7(3), pp. 141–152. <https://doi.org/10.1109/TITB.2003.813794>
- Karvelis, P.S., Fotiadis, D.I., Georgiou, I. and Syrrou, M. (2006). A Watershed Based Segmentation Method for Multispectral Chromosome Images Classification. *Conference Proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, 1, pp. 3009–3012. <https://doi.org/10.1109/IEMBS.2006.260682>
- Kayaalti, Ö., Aksebzeci, B.H., Karahan, İ.Ö., ... Asyali, M.H. (2014). Liver Fibrosis Staging Using CT Image Texture Analysis and Soft Computing. *Applied Soft Computing*, 25, pp. 399–413. <https://doi.org/10.1016/j.asoc.2014.08.065>
- Keltch, B., Lin, Y. and Bayrak, C. (2014). Comparison of AI Techniques for Prediction of Liver Fibrosis in Hepatitis Patients. *Journal of Medical Systems*, 38(8), p. 60. <https://doi.org/10.1007/s10916-014-0060-y>
- Ketteler, R. and Kriston-Vizi, J. (2016). High-Content Screening in Cell Biology. In *Encyclopedia of Cell Biology*. Elsevier, pp. 234–244. <https://doi.org/10.1016/B978-0-12-394447-4.40041-6>
- Kong, J., Cooper, L., Kurc, T., Brat, D. and Saltz, J. (2011). Towards Building Computerized Image Analysis Framework for Nucleus Discrimination in Microscopy Images of Diffuse Glioma. *Conference Proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, 2011, pp. 6605–8. <https://doi.org/10.1109/IEMBS.2011.6091629>
- Kononenko, I. and Kukar, M. (2007). *Machine Learning and Data Mining: Introduction to Principles and Algorithms*. Horwood Publishing Limited. Available at: <https://dash.harvard.edu/handle/1/3426318>
- Kuznetsov, Y.G., Chang, S.-C., Credaroli, A., Martiny, J. and McPherson, A. (2012). An Atomic Force Microscopy Investigation of Cyanophage Structure. *Micron (Oxford, England: 1993)*, 43(12), pp. 1336–42. <https://doi.org/10.1016/j.micron.2012.02.013>
- Kuznetsov, Y.G. and McPherson, A. (2011). Atomic Force Microscopy in Imaging of Viruses and Virus-Infected Cells. *Microbiology and Molecular Biology Reviews*, 75(2), pp. 268–285. <https://doi.org/10.1128/MMBR.00041-10>
- Leonard, A.P., Cameron, R.B., Speiser, J.L., ... Rohrer, B. (2015). Quantitative Analysis of Mitochondrial Morphology and Membrane Potential in Living Cells Using High-Content Imaging, Machine Learning, and Morphological Binning. *Biochimica et Biophysica Acta*, 1853(2), pp. 348–60. <https://doi.org/10.1016/j.bbamcr.2014.11.002>
- Li, K., Yin, J., Lu, Z., ... Liu, W. (2012). Multiclass Boosting SVM Using Different Texture Features in HEp-2 Cell Staining Pattern Classification. In *21st International Conference on Pattern Recognition*. IEEE, pp. 170–173. Available at: <http://ezproxy.uninorte.edu.co:2074/articleDetails.jsp?arnumber=6460099>
- Liao, M., Zhao, Y., Li, X., ... Zou, B. (2015). Automatic Segmentation for Cell Images Based on Bottleneck Detection and Ellipse Fitting. *Neurocomputing*. <https://doi.org/10.1016/j.neucom.2015.08.006>
- Lim, H.N., Mashor, M.Y. and Hassan, R. (2012). White Blood Cell Segmentation for Acute Leukemia Bone Marrow Images. In *2012 International Conference on Biomedical Engineering (ICoBE)*. IEEE, pp. 357–361. <https://doi.org/10.1109/ICoBE.2012.6179038>
- Lino, A., Rocha, A. and Sizo, A. (2016). Virtual Teaching and Learning Environments: Automatic Evaluation with Symbolic Regression. *Journal of Intelligent & Fuzzy Systems*, 31(4), pp.2061-2072. <https://doi.org/10.3233/JIFS-169045>

- Liu, Q. and Liu, X. (2013). Feature Extraction of Human Viruses Microscopic Images Using Gray Level Co-Occurrence Matrix. In *2013 International Conference on Computer Sciences and Applications*. IEEE, pp. 619–622. <https://doi.org/10.1109/CSA.2013.149>
- Liu, Y., Aickelin, U., Feyereisl, J. and Durrant, L.G. (2013). Wavelet Feature Extraction and Genetic Algorithm for Biomarker Detection in Colorectal Cancer Data. *Knowledge-Based Systems*, 37, pp. 502–514. <https://doi.org/10.1016/j.knsys.2012.09.011>
- Mao, Y., Liu, H., Ye, R., Shi, Y. and Song, Z. (2014). Detection and Segmentation of Virus Plaque Using HOG and SVM: Toward Automatic Plaque Assay. *Bio-Medical Materials and Engineering*, 24(6), pp. 3187–3198. <https://doi.org/10.3233/BME-141140>
- Matula, P., Kumar, A., Wörz, I., ... Rohr, K. (2009). Single-Cell-Based Image Analysis of High-Throughput Cell Array Screens for Quantification of Viral Infection. *Cytometry. Part A: The Journal of the International Society for Analytical Cytology*, 75(4), pp. 309–318. <https://doi.org/10.1002/cyto.a.20662>
- Mouelhi, A., Sayadi, M., Fnaiech, F., Mrad, K. and Romdhane, K. Ben. (2013). Automatic Image Segmentation of Nuclear Stained Breast Tissue Sections Using Color Active Contour Model and an Improved Watershed Method. *Biomedical Signal Processing and Control*, 8(5), pp. 421–436. <https://doi.org/10.1016/j.bspc.2013.04.003>
- Nanni, L. and Lumini, A. (2007). Ensemblator: An Ensemble of Classifiers for Reliable Classification of Biological Data. *Pattern Recognition Letters*, 28(5), pp. 622–630. <https://doi.org/10.1016/j.patrec.2006.10.012>
- Nosaka, R., Yasuhiro, O. and Kasuhiro, F. (2012). Feature Extraction Based on Co-Occurrence of Adjacent Local Binary Patterns. In Y.-S. Ho, ed. *Advances in Image and Video Technology*. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 82–91. <https://doi.org/10.1007/978-3-642-25346-1>
- Nowak, E., Jurie, F. and Triggs, B. (2006). Sampling Strategies for Bag-of-Features Image Classification. In A. Leonardis, H. Bischof, and A. Pinz, eds. *Computer Vision – ECCV 2006*. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 490–503. <https://doi.org/10.1007/11744085>
- Ojala, T., Pietikainen, M. and Maenpaa, T. (2002). Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), pp. 971–987. <https://doi.org/10.1109/TPAMI.2002.1017623>
- Ojansivu, V. and Heikkilä, J. (2008). Blur Insensitive Texture Classification Using Local Phase Quantization. In A. Elmoataz et al., eds. *Image and Signal Processing*. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-540-69905-7>
- Ong, H. and Chandran, V. (2005). Identification of Gastroenteric Viruses by Electron Microscopy Using Higher Order Spectral Features. *Journal of Clinical Virology: The Official Publication of the Pan American Society for Clinical Virology*, 34(3), pp. 195–206. <https://doi.org/10.1016/j.jcv.2005.04.001>
- Paci, M., Nanni, L., Lahti, A., ... Severi, S. (2013). Non-Binary Coding for Texture Descriptors in Sub-Cellular and Stem Cell Image Classification. *Current Bioinformatics*, 8(2), pp. 208–219
- Pudil, P., Novovičová, J. and Kittler, J. (1994). Floating Search Methods in Feature Selection. *Pattern Recognition Letters*, 15(11), pp. 1119–1125. [https://doi.org/10.1016/0167-8655\(94\)90127-9](https://doi.org/10.1016/0167-8655(94)90127-9)
- Sagonas, C., Marras, I., Kasampalidis, I., ... Karayannopoulou, G. (2013). FISH Image Analysis Using a Modified Radial Basis Function Network. *Biomedical Signal Processing and Control*, 8(1), pp. 30–40. <https://doi.org/10.1016/j.bspc.2012.05.001>
- dos Santos, F.L.C., Paci, M., Nanni, L., Brahmam, S. and Hyttinen, J. (2015). Computer Vision for Virus Image Classification. *Biosystems Engineering*, 138, pp. 11–22. <https://doi.org/10.1016/j.biosystemseng.2015.01.005>
- Seigel, T.A., Cocchi, M.N., Saliccioli, J., ... Donnino, M.W. (2012). Inadequacy of Temperature and White Blood Cell Count in Predicting Bacteremia in Patients with Suspected Infection. *The Journal of Emergency Medicine*, 42(3), pp. 254–9. <https://doi.org/10.1016/j.jemermed.2010.05.038>
- Shamir, L., Delaney, J.D., Orlov, N., Eckley, D.M. and Goldberg, I.G. (2010). Pattern Recognition Software and Techniques for Biological Image Analysis. *PLoS Computational Biology*, 6(11), p. e1000974. <https://doi.org/10.1371/journal.pcbi.1000974>
- Sonka, M., Hlavac, V. and Boyle, R. (2007). Image Processing, Analysis, and Machine Vision. Available at: <http://dl.acm.org/citation.cfm?id=1210103>
- Stoklasa, R., Majtner, T. and Svoboda, D. (2014). Efficient K-NN Based HEp-2 Cells Classifier. *Pattern Recognition*, 47(7), pp. 2409–2418. <https://doi.org/10.1016/j.patcog.2013.09.021>
- Tan, X. and Triggs, B. (2010). Enhanced Local Texture Feature Sets for Face Recognition under Difficult Lighting Conditions. *IEEE Transactions on Image Processing*, 19(6), pp. 1635–1650. <https://doi.org/10.1109/TIP.2010.2042645>
- Tarnawski, W., Kurtcuoglu, V., Lorek, P., ... Ferrari, A. (2013). A Robust Algorithm for Segmenting and Tracking Clustered Cells in Time-Lapse Fluorescent Microscopy. *IEEE Journal of Biomedical and Health Informatics*, 17(4), pp. 862–9. <https://doi.org/10.1109/JBHI.2013.2262233>

- Tonti, S., Di Cataldo, S., Bottino, A. and Ficarra, E. (2015). An Automated Approach to the Segmentation of HEp-2 Cells for the Indirect Immunofluorescence ANA Test. *Computerized Medical Imaging and Graphics: The Official Journal of the Computerized Medical Imaging Society*, 40, pp. 62–69. <https://doi.org/10.1016/j.compmedimag.2014.12.005>
- Wang, J., Zhou, X., Li, F. and Wong, S. (2006). Classify Cellular Phenotype in High-Throughput Fluorescence Microcopy Images for RNAi Genome-Wide Screening. In *2006 IEEE/NLM Life Science Systems and Applications Workshop*. IEEE, pp. 1–2. <https://doi.org/10.1109/LSSA.2006.250404>
- Woźniak, M., Graña, M. and Corchado, E. (2014). A Survey of Multiple Classifier Systems as Hybrid Systems. *Information Fusion*, 16, pp. 3–17. <https://doi.org/10.1016/j.inffus.2013.04.006>
- Yi, F. and Moon, I. (2012). Image Segmentation: A Survey of Graph-Cut Methods. In *2012 International Conference on Systems and Informatics (ICSAI2012)*. IEEE, pp. 1936–1941. <https://doi.org/10.1109/ICSAI.2012.6223428>
- Ylioinas, J., Hadid, A., Guo, Y. and Pietikäinen, M. (2013). Efficient Image Appearance Description Using Dense Sampling Based Local Binary Patterns. In K. M. Lee et al., eds. *Computer Vision – ACCV 2012*. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 375–388. <https://doi.org/10.1007/978-3-642-37431-9>
- Zhang, J., Wang, Y. and Shi, X. (2009). An Improved Graph Cut Segmentation Method for Cervical Lymph Nodes on Sonograms and Its Relationship with Node's Shape Assessment. *Computerized Medical Imaging and Graphics: The Official Journal of the Computerized Medical Imaging Society*, 33(8), pp. 602–607. <https://doi.org/10.1016/j.compmedimag.2009.06.002>