A Spiral-Based Methodology Applied to Cell Image Analysis

Margarita R. Gamarra Acosta[a], Evert De los Ríos Trujillo[b] & José Escorcia-Gutierrez[c]

[a] *Ingeniería Electrónica, Politécnico de la Costa Atlántica, Barranquilla, Colombia.*
mgamarraa@pca.edu.co
[b] *Ingeniería Electrónica, Universidad del Magdalena, Santa Marta, Colombia.*
edelosrios@unimagdalena.edu.co
[c] *Centro de Investigaciones, Escuela Naval de Suboficiales A.R.C. "Barranquilla", Barranquilla, Colombia.*
jose.escorciag@armada.mil.co

Como citar: Gamarra Acosta, M.R., De los Ríos Trujillo, E. & Escorcia-Gutierrez, J. (2019). A spiral based methodology applied to cell image analysis. Revista Sextante, 20, pp. 4-11, 2019.

Abstract

The advances in technology, microscopy and computing have allowed the development of new fields in cell image analysis. However, the usability of these platforms is adequate to expert users only. Many software tools are oriented to expert users in image processing, likewise the use of bioinformatics require a basic knowledge in programming. The development of research in cell imaging requires the joint work of computer Scientifics and biologist. In this paper we present a methodology to develop a software solution applied to the analysis of cell images.

Keywords: Bioinformatics; Cell image; Image processing; Software methodology.

Una metodología basada en espiral aplicada al análisis de células en una imagen

Resumen

Los avances en tecnología, microscopía y computación han permitido el desarrollo de nuevos campos en el análisis de imágenes celulares. Sin embargo, la usabilidad de estas plataformas es adecuada solo para usuarios expertos. Muchas herramientas software están orientadas a usuarios expertos en el procesamiento de imágenes y así mismo el uso de herramientas bioinformáticas requiere un conocimiento básico en programación. El desarrollo de investigaciones en imágenes celulares requiere el trabajo conjunto de biólogos y de expertos en computación. En este artículo se presenta una metodología para desarrollar una solución de software aplicada al análisis de imágenes celulares.

Palabras claves: Bioinformática; Imagen de célula; Procesamiento de imágenes; Metodología de software.

1. Introducción

Current researches in biology include both quantitative and qualitative analysis, which is important for clinical applications (e.g., diagnosis and treatment of disease, vaccine development, etc.) and biological research (e.g., to understand the cell behavior, virus infection). In these studies, the amount of information and metadata contained in a single sample is large. In the case of cell images, there are millions of cells in only one image obtained by different microscopy technologies, and then in the presence of millions of cells, the human visual classification becomes infeasible.

The introduction of new models, measurements and methods has produced an amount of data using image-based evidence (Hanchuan, 2018). However the big data and the extracted information of an image are difficult to organize, search, process and analyze. As a natural extension of the existing biomedical image analysis field, the bio-image informatics develops and uses various image data analysis and informatics techniques to extract, compare, search and manage the biological knowledge of the respective images (Hanchuan, 2018).

There are many software tools useful to the researchers, which include several image processing techniques and data analysis. In the cell image case some problems arise: techniques are limited and not work well in all situations (e.g., cells close together are difficult to segment). The parameters selection for the image processing algorithms, are unknown for biologist and it is necessary create a software program to improve the established techniques (it requires an expert in image processing). In addition, although the stages in the image processing are well defined (acquisition, segmentation, feature extraction and classification), implementing a software development methodology is necessary in order to create a framework and improve the results according to the specific issue.

Our primary aim in this paper is to propose a spiral-based methodology which meets the needs of biologists, scientists and engineers to extract useful and novel knowledge from cell images, using enhanced algorithms. The bioinformatics research scenario we chose is cell image analysis, with a case-study with a specific application to analyze the infected cells by Respiratory Syncytial Virus (RSV). This field is widely studied because of the information obtained of a cell image is highly significant in the disease research.

A spiral-based methodology to software development is adequate for large projects, like cell image informatics, and its flexibility allows changes to be implemented at several stages of the project. This methodology allows incorporate the step of the waterfall method to the stages of digital image processing. The main advantage of this methodology is that the client (biologists, scientists or laboratory technicians), who will be involved in the development of each segment, retains control over the direction and implementation of the project (SQA, 2016). Furthermore, the proposed spiral-based methodology for software development in cell imaging could be used in other image processing projects.

We have organized this paper in order to explain the proposed methodology. In Section 2, we present some currently available platforms that offer similar services in cell image analysis. Likewise, we relate other published works and discuss these in relation to our proposal. In Section 3 we describe the software development methodology, including the phases, artifacts and roles involved in the project and we discuss some details about its implementation. In Section 4 we expose an implementation of the proposed methodology applied to the development of a platform for cell image analysis. Finally, we present the conclusions and future work.

2. Related Works

In recent years the management of big data and cloud computing has increased the researches that use big data processing. Many biological images management tools have been built for specialized domains. We will focus on the platforms and software related with cell image analysis.

The following software are widely used for cell image processing:

CellProfiler (Carpenter, y otros, 2006) is free open-source software designed to enable biologists without training in computer vision or programming

to quantitatively measure phenotypes from thousands of images automatically.

ImageJ (Abràmoff, Magalhães, & Ram, 2004) is a Java based cross-platform tool for biomedical image processing and measurement. A number of image analysis toolboxes such as fluorophore tracking, filament detection, etc., were developed by various groups.

ITK (Yoo, et al., 2002) is an open-source, cross-platform system that provides developers with an extensive suite of software tools for image analysis. ITK employs leading-edge algorithms for registering and segmenting multidimensional data.

Some Platform-as-a-Service PaaS has been developed, integrating cell image processing with web services:

The UCSB Bisque (Kvilekval, Fedorov, Obara, Singh, & Manjunath, 2010) system provides an integrated online environment for users to upload, search, edit and annotate images. Although it is not specialized in cell images, it includes a few analysis and visualization modules for this field.

Authors in (Pelet, Dechant, Lee, Van, & Peter, 2012) have developed an image analysis platform named YeastQuant to simplify data extraction by offering an integrated method to turn time-lapse movies into single cell measurements. The database is connected to the engineering software Matlab, which allows extracting the desired information by automatically segmenting and quantifying the microscopy images.

The research developed in (Noura, Al-Otaibi, & Amin, 2011) suggests the use of Service Oriented Architecture (SOA) to integrate biological data from different data sources. This work shows that SOA solves the problems of facing integration process and accessing biological data in easier way. The Microsoft .Net Framework was used to implement proposed architecture.

Authors in (Juan Carlos, Francisco, Vicente, & M. Carmen, 2014) present a web services-based platform focused on the cell counting problem. Using OpenCF, a web services development framework, authors integrate in a single platform services oriented to image processing and classifying, cell counting based on a set of parameters, and data post-processing (plot generation, datasheets, etc.). A GUI added to the platform helps to launch jobs with image sets, and the execution of different tasks from a web service-based client.

3. Progressive Spiral Methodology

This methodology reflects the underlying concept that each cycle involves a progression that addresses the same sequence of steps, for each portion of the product and for each of its levels of elaboration, from an overall concept of operation document down to the coding of each individual program (Boehm, 1988). An important feature of the spiral model is that each cycle is completed by a review involving the primary people or organizations concerned with the product. Another advantage is the facility to complete the process in any stage, this means that the final objective is not the classification; it could be only improving quality, segmentation, cells counting, extracting features to realize statistical analysis or finally a classification. Additionally, the applied methodology is progressive due to the process is developed gradually or in stages, proceeding step by step.

The main stages in a digital image processing are:

Acquisition: Before any image processing, an image must be captured by a device and converted into a manageable entity. The image acquisition process consists of three steps: energy reflected from the object of interest, an optical system which focuses the energy and finally a sensor which measures the amount of energy (Asit Kumar & Soumika, 2016).

Preprocessing: this stage involves other concepts (Gonzales, 2009):

Image enhancement: Basically, the idea behind enhancement techniques is to bring out details that are obscured, or simply to highlight certain features of interest in an image.

Image restoration: Unlike enhancement, which is subjective, image restoration is objective, in the sense that restoration techniques tend to be based on

mathematical or probabilistic models of image degradation.

Segmentation procedures partition an image into its constituent parts or objects. In general, autonomous segmentation is one of the most difficult tasks in digital image processing. Compared with the algorithms for common image processing, the ones used for medical images require more concrete application background (Zhen, Manuel, R, S, & M, 2009) and different algorithms have been used for medical image segmentation (Ma, Tavares, Jorge, & Mascarenhas, 2010; Danilo Samuel, Aledir Silveira, João, & R.S, 2016). A rugged segmentation procedure brings the process a long way toward successful solution of imaging problems that require objects to be identified individually.

- *Feature extraction* deals with extracting attributes that result in some quantitative information of interest or are basic for differentiating one class of objects from another (Gonzales, 2009; Roberta B, João P, Aledir S, & João Manuel, 2018).

- *Classification* is the process that assigns a label to an object based on its descriptors (Gonzales, 2009). With the aim to select the appropriated classifier, two or three classifiers should be tested. Features like computational cost, hit rate and evaluation time are compared between the classifiers.

We have adapted the spiral model to our proposal, merging the spiral form with the stages for image processing.

In the center of the spiral is the core of the project, which includes the definition of the main goal and a global vision of the entire plan. This preceding phase can avoid delays or going back to a previous stage during the project. This core phase can be implemented in one or two weeks.

Then, some questions must be answered in this core phase before to start the project:

-What is the main objective?

-What are the restrictions and requirements?

-Is there any previous work related to this project?

-What techniques are commonly used?

-Are the resources available during all stages?

The proposed methodology is represented in [Figure 1](), and it is composed by four main phases:

A. Analysis: in this phase the development team exposes the objectives, alternatives and constrains. Each stage of the image processing has a different analysis and it is important the discussion between the development team and the expert biologists. This phase could be developed during one week.

B. Evaluation: in this phase the possibilities and constrains exposed in the previous phase are evaluated, the risks are identified, and the suitable solution is selected. Likewise, the expert opinion is significant. This phase requires one or two weeks.

C. Modeling: in this phase the objects of study (images, pixels, features, and vectors) are the inputs to the model, which is a mathematical representation to perform the processing in each stage (acquisition, preprocessing, feature extraction and classification). The verification is important in this phase because the modeling leads to the implementation. This phase takes longer than the previous phases because the research and modeling could take about one or two weeks, depending on the research.

D. Development: in this phase the algorithms or the techniques are implemented. In many cases a machine with a high processing capacity is required, because of the size and the amount of the images to process and the complex mathematical operations. This is a strong reason to employ cloud computing for cell images processing. Once again, the verification and assessing of the health or biology expert is essential in this phase. This phase is the longest because it implicates the development of algorithm and test.

According to the CMMI® (Capability Maturity Model® Integration), the verification and validation process must be included in the software development projects. The purpose of Verification is to ensure that selected work products meet their specified requirements (Institute, 2010).
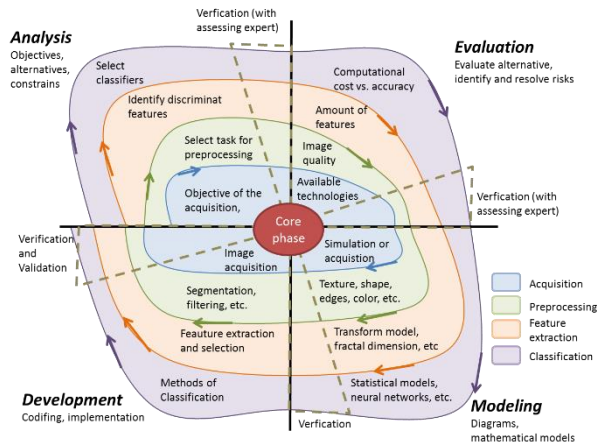
Figura 1. Progressive Spiral Methodology (PSM).
Source: The authors.

Verification is inherently an incremental process because it occurs throughout the development of the product and work products, beginning with verification of requirements, progressing through the verification of evolving work products, and culminating in the verification of the completed product. The purpose of Validation is to demonstrate that a product or product component fulfills its intended use when placed in its intended environment (Institute, 2010).

In our proposal, for each phase, an expert verification and assessing process is added, it aims to guarantee the results and to involve the expert judgment. In the modeling phase, only a process of verification is necessary, because the mathematical methods used for processing do not involve health or biology experts. Furthermore, when each stage has finished a validation process is performed in order to guarantee compliance of objectives.

4. Methodology applied to RSV análisis

The proposed methodology is applied to the study of human cells infected with RSV (Respiratory Syncytial Virus). This research aims to develop a morphological characterization and statistical analysis of image cell under laboratory conditions, to facilitate the identification of population growth that contribute to the understanding of mechanisms of entry of RSV to the body human. The used image database corresponds to Human epithelial type 2 (HEp-2) cells infected with Respiratory Syncytial Virus (RSV). The set of images contains different density of cells seeded, which allows a more complete analysis of the viral infection.

The four phases described in Section 3 are applied to the RSV research. The spiral-based methodology is applied in each stage of the project, as shown in Figure 2. The global vision at the start of the project and the recommended techniques give a benchmark for the development team. The final selection is defined by the performance of the algorithms.

Core Phase

In this initial phase of the project the project manager and the development team had a meeting with the experts in virology: a medical virologist, a biologist, and the laboratory technicians. They exposed the goals, the restrictions (equipment, devices, cell samples, amount and quality of the images, time and available budget) and requirements (security in the platform and data management, image quality, big-data processing, usability of the platform, accuracy for the results, availability database, compatibility). In this stage the researches give a global approach about the development of the project and it is necessary a review of the state-of-the-art.

An initial guide to start the project is based on proposed questions in Section 3.

-What is the main objective? The main objective of the image processing is to classify the cell in two classes: infected and non-infected. Although this is not the general objective of the research, the results of the image processing contribute to the posterior analysis about entry mechanisms of RSV to the human cells.

-Is there any previous work related to this project? There are several related works about platforms and software development for cell image analysis (see Section 2). Additionally, some cell image processing research has been developed in recent years.

-What techniques are commonly used? Innovative and well-known techniques have been implemented in each stage of the cell image processing:

Acquisition: Fluorescence microscopy, confocal microscopy, TEM, AFM.

Preprocessing (segmentation): Thresholding, Region-based segmentation, Edge-based segmentation, Energy-based segmentation, clustering.

Feature extraction: Size and structure, Texture, Color.

Classification: Statistical and classical, Artificial intelligence, Boosting.

-What are the restrictions and requirements? Based on the requirements, restrictions, technologies and methods, the development team selects the database and the processing algorithms.

Analysis, Evaluation, Modeling and development Phases

The four phases described in Section 3 are applied to the RSV research. In Figure 2 the spiral-based methodology is applied in each stage of the project. During the project, the techniques in the processing could be changed, but the global vision at the start of the project and the recommended techniques gives a benchmark for the development team. The final selection is defined by the performance of the algorithms.

In Table 1 the matrix of artefacts, techniques and roles is applied to the RSV project, the preprocessing, features extraction and classification stages.

These matrixes are related with the proposed phases for the spiral-based methodology shown in Figure 2. The implementation of the four phases, with the verification and validation phases, allowed an organized progress of the image acquisition and preprocessing. Additionally, involving experts in each phase generated satisfactory results.

In Figure 3 an example of the acquired images is shown. For this stage, a fluorescence microscope (Reference: Axio Observer, Zeiss) was used and the goal was reached (visualize infected and non-infected cells). Figure 3 c) shows the results of the segmentation algorithm applied.
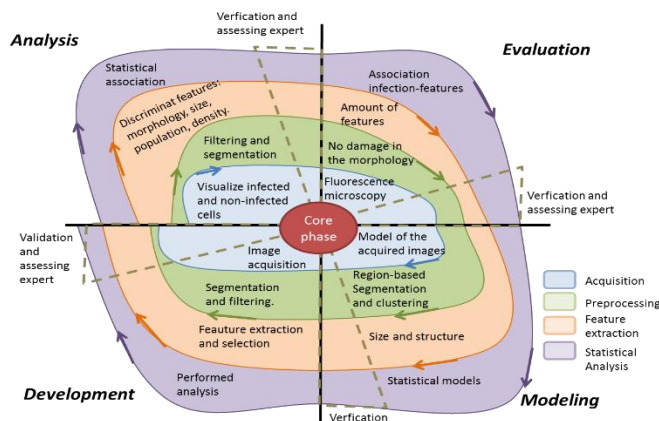

Figure 2. Spiral-based methodology applied to RSV project.
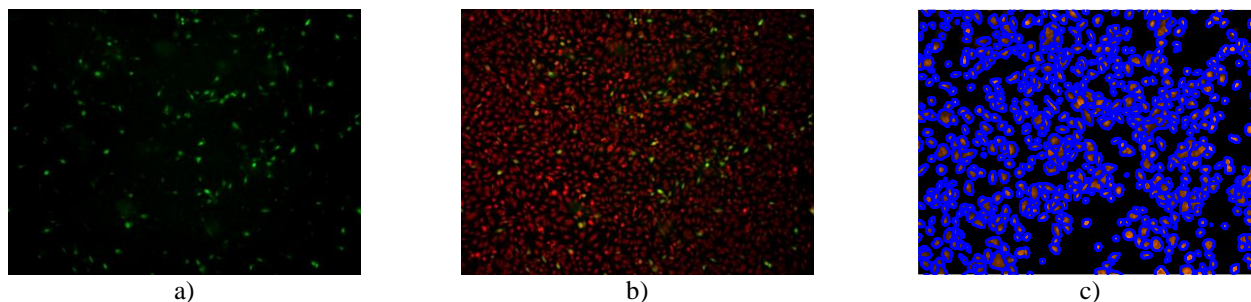**Source:** The authors.

Figure 3. a) Infected cells in green channel. b) Infected and non-infected cells in green and red channels. c) Segmented cell.
Source: The authors.

Table 1. Matrix applied to RSV project

Phase	Stage	Artifacts	Techniques	Roles
Analysis	Acquisition	Document about the devices and technologies to use.	Meetings with the manager and experts (biologists, laboratory technicians).	Project Manager, development team, experts.
	Preprocessing	Document about segmentation techniques.	Meetings with experts (biologists, laboratory technicians).	Development team, experts.
	Extraction	List with the potentials features to extract.	Analysis of the images: quality, elements, size, colors, etc.	Development team, experts.
	Statistical analysis	Selection of statistical models and test.	Analysis of the data type (binary) and the appropriated information.	Development team, experts.
Evaluation	Acquisition	Selection of the best option: buy or rent devices, simulate images or use a database.	Evaluate the budget and the accessibility to the devices or images.	Project Manager, development team, experts
	Preprocessing	Document with requirements about preserve morphology and	Review of the state of the art and previous works.	Development team, experts.
	Extraction	Selection of features to extract.	Review of the state of the art and previous works.	Development team, experts.
	Statistical analysis	Report about metrics, indicators and confidence intervals.	Review of the state of the art and previous works.	Development team, experts.
Modeling	Acquisition	Image model.	Simulation techniques or methods of acquisition.	Development team.
	Preprocessing	Mathematical models applied in segmentation.	Modeling of preprocessing operations.	Development team.
	Extraction	Representation of the extracted features like a vector or matrix.	Dimension changes, operations on pixels or the image, etc.	Development team.
	Statistical analysis	Documentation about statistical test and models.	Statistical modeling.	Development team.
Development	Acquisition	Digital images.	Microscopy, simulation.	Development team, experts.
	Preprocessing	Enhancement and Segmentation.	Filtering, color changes, etc.	Development team, experts.
	Extraction	Vector of features.	Application of extraction methods.	Development team, experts.
	Statistical analysis	Statistical validations, indicators and confidence intervals.	Implementation of statistical modeling.	Development team, experts.

Source: The authors.

5. Conclusions

The development of new technologies in microscopy has improved the study of cells using 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. In this article we have proposed a methodology for development of projects in cell image analysis.

This spiral-based methodology was applied to the RSV project and the results in the acquisition, segmentation, feature extraction and statistical analysis were satisfactory. For each of these stages the application of the four phases proposed in the methodology: analysis, evaluation, modeling, and development.

References

Abràmoff, M., Magalhães, P. J., & Ram, S. J. (2004). Image Processing with ImageJ. Biophotonics, 1-7.

Asit Kumar, D., & Soumika, M. (2016). Information Photonics: Fundamentals, Technologies, and Applications. B/W

Illustrations. Retrieved from Image Acquisition: http://www.springer.com/cda/content/document/cda_downloaddocument/9781447125020-c2.pdf%3FSGWID%3D0-0-45-1284937-p174272879+&cd=2&hl=es&ct=clnk&gl=co

Boehm, B. W. (1988). *A spiral model of software development and enhancement. Computer,* 61-72.

Carpenter, A., Jones, T., Lamprecht, M. R., Clarke, C., Kang, I. H., Friman, O., . . . Sabatini, D. M. (2006). *CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biology,* 1-11.

Danilo Samuel, J., Aledir Silveira, P., João, M., & R.S, T. (2016). *A review of computational methods applied for identification and quantification of atherosclerotic plaques in images. Expert Systems with Applications,* 1-14.

Gonzales, R. (2009). *Digital Image Processing. Person Education,* 1-8.

Hanchuan, P. (2018). *Bioimage informatics: a new area of engineering biology. Bioinformatics,* 1827-1836.

Institute, C. (2010). *CMMI for Development, Version 1.3. CMMI Institute.*

Juan Carlos, C., Francisco, A., Vicente, B., & M. Carmen, R. (2014). *Web Services Based Platform for the Cell Counting Problem. European Conference on Parallel Processing,* 83-92.

Kvilekval, K., Fedorov, D., Obara, B., Singh, A., & Manjunath, B. (2010). *Bisque: a platform for bioimage analysis and management. Bioinformatics,* 544-552.

Ma, Z., Tavares, J., Jorge, R., & Mascarenhas, T. (2010). *A review of algorithms for medical image segmentation and their applications to the female pelvic cavity. Comput Methods Biomech Biomed Engin,* 235-246.

Noura, M., Al-Otaibi, & Amin, Y. N. (2011). *Biological Data Integration using SOA. International Journal of Computer, Electrical, Automation, Control and Information Engineering,* 1-6.

Pelet, S., Dechant, R., Lee, S., Van, D., & Peter, M. (2012). *An integrated image analysis platform to quantify signal transduction in single cells. IntegrBiol,* 1274-1282.

Roberta B, O., João P, P., Aledir S, P., & João Manuel, R. S. (2018). *Computational methods for pigmented skin lesion classification in images: review and future trends. Neural Computing and Applications,* 613-636.

SQA. (2016, Abril 7). *Advantages of the Spiral Model. Retrieved from Scottich Qualifications Authority:* https://www.sqa.org.uk/e-learning/SDM01CD/page_10.htm

Yoo, T., Ackerman, M., Lorensen, W., Schroeder, W., Chalana, V., Aylward, S., . . . R. (2002). *Engineering and algorithm design for an image processing Api: a technical report on ITK--the Insight Toolkit. Stud Health Technol Inform,* 586-592.

Zhen, M., Manuel, J., R, S, T. R., & M, N. J. (2009). *A review on the current segmentation algorithms for medical images. Porto, Portugal: Faculty of Engineering, University of Porto.*