

# **Automation of the process of vectorization in areas of interest and buffer variable on satellite images**

**Éric BELTRÃO and Rudiney PEREIRA, Brazil**

**Key words:** python, automation, geoprocessing, vectorization, geographical information system.

## **SUMMARY**

The constant improvement in computer application development techniques, combined with the technological advancement of electronic devices that support such systems, comes with the most contributing significantly different areas of knowledge, adding to these, often, greater accuracy and agility in their processes. Based on this context, and returning our looks for areas of geographic information System (GIS), and Digital image processing, the development of two programs that serve as a proposal of technological innovation as regards the vectorization of areas of interest on satellite images, and creating the buffer variable on the same, this variation that follows the characteristics of width of the area considered. Both programs were developed in the Python programming language and has as final product a shape file containing the vectorization of all areas found in a satellite image by means of classification or filtering. Also presents itself in this work the manipulation of files generated, in order to guide the actions necessary to use, in the best way, the product of these programs, thus helping the user of GIS computer systems for the performance and accuracy of their work.

## **RESUMO**

O constante aperfeiçoamento nas técnicas de desenvolvimento de aplicativos computacionais, associado ao avanço tecnológico dos dispositivos eletrônicos que suportam tais sistemas, vem contribuindo significativamente com as mais diversas áreas do conhecimento, agregando a estas, muitas vezes, maior precisão e agilidade em seus processos. Com base neste contexto, e voltando nossos olhares para as áreas de Sistema de Informações Geográficas (SIG), e Processamento Digital de Imagens, realizou-se o desenvolvimento de um programa que servisse como proposta de inovação tecnológica no que diz respeito à vetorização de áreas de interesse sobre imagens de satélite. Este programa foi desenvolvido na linguagem de programação Python e tem como produto final um arquivo shape contendo a vetorização de todas as áreas encontradas em uma imagem de satélite por meio de classificação ou filtragem desta. Também se apresenta neste trabalho a manipulação dos arquivos gerados, a fim de nortear as ações necessárias para a utilização deste tecnologia, da melhor forma, o produto desse programa, desta maneira auxiliando o usuário dos sistemas computacionais de SIG no se refere ao desempenho e precisão do seu trabalho.

**Palavras-Chave:** python; automação; geoprocessamento; vetorização; sistema de informações geográficas.

# Automation of the process of vectorization in areas of interest and buffer variable on satellite images

Éric BELTRÃO and Rudiney PEREIRA, Brazil

## 1. INTRODUCTION

Every passing day, the computing systems are gaining more notoriety in our lives, modifying business dynamics, the relation between the educational institutions and their students and even the way the society interact.

It is not difficult to find individuals who are resistant to the technological advance, by arguing that computers have emerged to take jobs from other individuals, not realizing that if technological advance reduced or extinguished the number of workers in certain sectors of the economy, on the other hand this technology has produced new jobs and even created new areas of work that would not exist without such technology.

The advent of computing systems, mobile devices and modern communication systems, generated a demand for qualified labor in these areas, keeping in mind we are not talking just of conventional computers, but any device that has an embedded system. On this side, we can discuss the merits of many new professions, activities and jobs the information technology has produced, in the manufacturing of hardware, Software Development, selling and maintenance of these technologies, not counting with new paradigms that have emerged, as virtual companies and others. While recognizing the benefits brought by technological advancement to the present day, we cannot foster the illusion that computers are the solution for the problems of mankind. As Laurent Gasser would say (apud PORTAL REGIONAL, 2010), “Computers do not solve problems, they implement solutions”. Taking this idea as a basis, we should see this equipments as one more tool in our range of technical possibilities and different shapes for revolution of problems.

According to the French teacher (MAUBORGNE, 2010), that focuses on discipline of strategy, not every innovation is considered good, because there is a need to be useful and meet the demand of users. We can apply the same respect to computerized systems, since, not all launches, or new functionalities proposals will bring real advantages to its users, making these continue with the solutions previously used.

Geomatics is consolidating itself as an interdisciplinary field and has the multiple usability as a big deal, this is because, the use of geospatial information has been disseminated with great speed in today's society.

Nowadays, it becomes unquestionable the popularization of geospatial services, such as, online maps, where travel routes can be traced, specific places can be found, whether commercial, leisure or hosting. Another point, we cannot overlook the fact that, with technological advance, mobile devices started to gain spatial functionality, whether a system that communicates with a GPS via Bluetooth or, a device that already has embedded such provision, contributing to the popularization of service in this area.

Among the activities, typically performed on a computerized GIS is the representation of the real world. This representation can be made in two ways: by raster images (German term that refers to the grid or raster data), or by vectorization (vector data such as points, lines and areas). Vector data are represented by points, lines and polygons, that are ways of showing respectively a particular coordinated a path or way, and finally, the demarcation of an area (MIRANDA, 2006).

Among the technologies involved in GIS, is the digital image processing. Unlike techniques

that work with vector data, this technology has its focus on raster data, which contain a numerical representation pixel by pixel of the generated image whether monochrome or colorful. The idea of interaction between the methodologies for digital image processing and vectorization, together with the aim of optimize performance and accuracy in defining the areas of interest, were crucial points of motivation for the development of this work (MIRANDA, 2010).

Throughout this article, we describe the utilization of geographic information System (GIS), of digital image processing, as well as is demonstrated the use of a prototype (something still in stage of testing or development) of software developed as a proposed technological innovation in the field of GIS. This software was developed with the intention of automate procedures that, until now, has been performed by users of computer programs in GIS, more specifically related to the delimitation of areas of interest through vectorization.

## 2. MATERIALS AND METHODS

### 2.1 Test Image

For the development of prototype software, it was not necessary to use an image in full, opting to perform the clipping of an area located in the upper left quadrant of the image named "LANDSAT\_5\_TM\_20090411\_223\_081\_L2\_BAND4" obtained the image from the catalog site of the National Institute for Space Research (INPE), dated de April 11<sup>th</sup> 2009, without cloud cover and belonging to the band 4 of TM sensor. This area contains all the necessary features for the development of the research, is represented by clipping UTM XY coordinates of upper left corner (687236, 6707269), lower left (687236, 6694754), upper right (705542, 6707269) and bottom right (705,542, 6,694,754).

It should be emphasize that we are not taking into account the spatial resolution of the image, because it is a prototype, and it must work in different resolutions, we adopted a system variable responsible for controlling the functionality of the software, according to the resolution. To this variable was admitted different resolutions, thus to analyze system behavior.

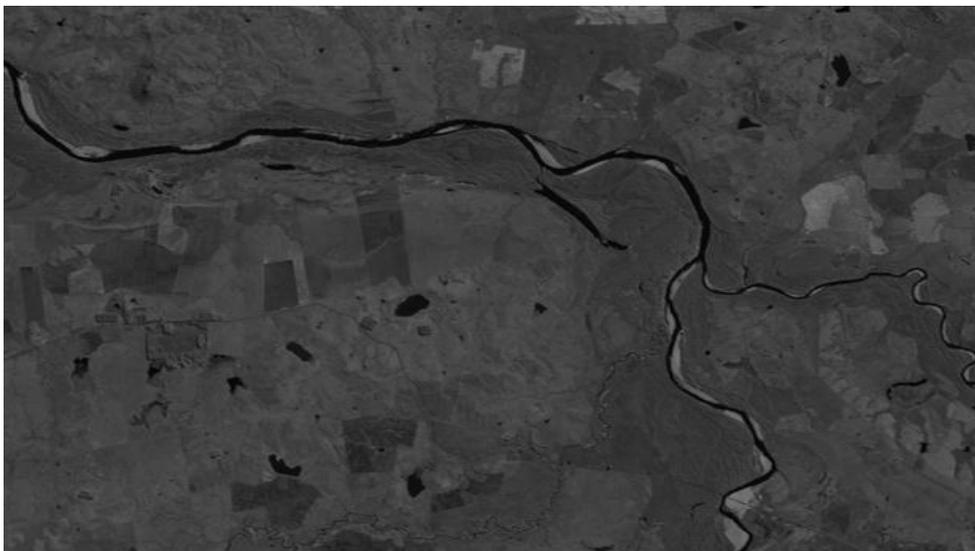


Image 1 - Clipping image band 4 TM sensor - INPE (2011)

## 2.2 Python Programming Language

The programming language used was Python, because it has gained more space each Day in the international developing market of computing devices. Some of its features will be described in the next paragraphs in order to understand its specificities and its wide acceptance in the global stage. (LUNDH, 2012).

First, it should be highlighted that Python is an open source software, allowing interaction of the developer community with it reporting problems, proposing improvements, causing it to continue evolving at a rapid pace. Python is distributed through a license compatible with the General Public License (GPL), but with greater flexibility, allowing it to be used even for proprietary products. (BORGES, 2010).

## 2.3 Filter by frequency

With the help of the features of Python we used a filtering technique which operates in a frequency band of the image. This type of filter falls into the category of bandpass filters that acts highlighting a range of values of interest, and rejecting all other frequencies outside.

The image 2 presents a similar code to that used to perform filtering on the image shown in figure 1. This code shows the importation of PIL library that works specifically with images represented by line 1. The opening of the image in line 3 of the code. The filtering itself, in line 5, which was adopted as the frequency range of interest of less than 30 images , being understood that these would be elevated to the value 255 represents white shade in the image and, lastly, its display and saving in the following lines.

```
1 import Image
2
3 imagem = Image.open("entrada.tif")
4
5 img = imagem.point(lambda i: i < 30 and 255)
6
7 img.show()
8
9 img.save("saida.png")
```

Image 2 - Similar code to that used to achieve filtering.

The result of using the filter described previously is showed in the image 3, in which we can clearly identify the pixels representing the water body as highlighted pixels, thereby reducing the complexity of identifying the edges of a body of water between so many variations on the original pixel.



Image 3 - Product filtering.

## 2.4 Developing the Software

By implementing the computer program, we made a scan in the image that took into consideration the matrix characteristics that are found, both the original image and the product generated by filtering it. Therefore, the analysis performed looking for areas of interest occurs tracing the image in a direction from left to right, analyzing all the columns from the first line, and in order from the top down, because the process performed on the first line ends up repeating in consecutive lines until it reaches the end of the picture.

When we find a pixel of interest, by scanning described before, the software calls a secondary feature denominated “delimitada ()” (delimit). Functionality known among developers of several programming languages as 'method', and it is used to limit the specificities of the development , making it somehow more flexible because they may have their call at any time and from any part of the code.

In the act of calling the method “delimita()”, this should receive the coordinates (x, y) of the first pixel of interest found, on the basis of which method is intended to record the edges of the area of interest to which the pixel belongs, and it is performed in a clockwise direction until you find the starting pixel.

After reaching the goal of storing all the points that form the polygon on the image, we used a Python class called shapefile in its version 1.1.3. With the assistance of this class, it was possible to create the shape file that later is shown on the test image in figure 4.

It is necessary to explain that, in the first tests conducted, we adopted as the default store as coordinate reference, just that represents the upper left corner of each pixel. This way, obviously, the area of interest has not been completely covered by the polygon generated via the prototype of vectorization, then trace adjustments have been made so that the whole area was properly covered by the polygon generated effectively.

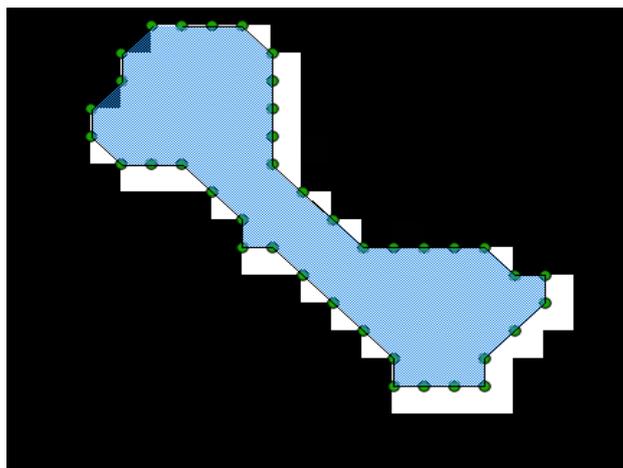


Image 4 – Shape file on test image.

### 3. RESULTS AND DISCUSSIONS

#### 3.1 Optimization of time and accuracy

Considering the tests performed with the software developed on the image represented by Figure 3, that is the product of filtering the image represented by Figure 1, we find excellent results in respect to the time of vectorization and accuracy thereof.

Regarding to the runtime of the program, firstly one has to understand the characteristics of the materials involved the image used had dimensions of 617 pixels wide by 424 pixels high, totaling 261,608 (two hundred and sixty-one thousand six hundred and eight) pixels to be analyzed. The basic configuration of the computer used: 2.0 Centrino processor CoreTM2 Duo, 3 GB RAM 667, and with 3 M cache.

Based on the features, mentioned previously, we obtained about 9 seconds of time to analyze the image and create the shape file by the developed prototype. This shapefile was composed of 131 polygons representing water bodies, and to represent each of them it was necessary four or more coordinates of the image, considering it should represent conversion from raster to vector data, accounting for a total of 7249 (seven thousand two hundred forty-nine) coordinates.

The scanning areas of interest is of utmost importance to the quality of the final product in GIS, however, if applied to large areas, it becomes relatively exhaustive and time-consuming for any professional. So, it is understandable the relentless search for improvement for the existing systems, and implementation of new products e and automation modules in GIS processes.

In the tests performed, the developed prototype demonstrated scan areas of interest with 100% coverage of these, following the edge of the pixels belonging to the same faithfully. In this case, the big difference in quality comes from the classification adopted for recognition of areas of interest; the better the rating, the better the scan.

#### 3.2 Manipulating shapefile with QGIS

Manipulating the shapefile generated automatically can be accomplished by various computational systems GIS. Nevertheless during the tests we used the Quantum GIS software 1.4.0. With the aid of this system we demonstrated the generated file with its 131 polygons as seen in Image 5.

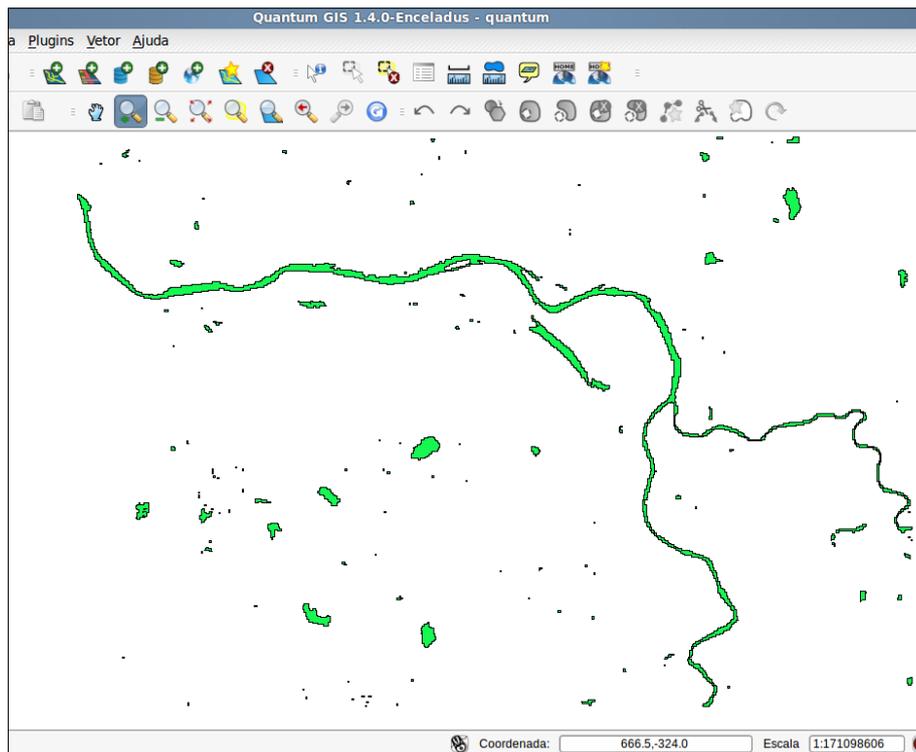


Image 5 –Shapefile resulting from vectorization on the test area.

In order to show the delimitation of APP on the banks of the water body based on the file that was generated automatically, there was the need to divide the features (polygons) existent, because the rules that apply to rivers and streams are not the same apply to lakes, ponds and reservoirs, natural or artificial.

For this purpose, it is used the functionality 'select features', using this functionality, it was possible to select all the features and remove the selection of one feature that represents the watercourse.

Containing only the lakes, ponds and water reservoirs selected, we made the use of a second software functionality of Quantum GIS named save selection as shapefile, which can be found in the menu Layer and enables the creation of a new shapefile containing only the previously selected features.

The same way, it is possible to create a shapefile containing only the features that represent waterways, and generalize the width of these, and even to perform divisions regarding the existence of files divided by the width of rivers to which they relate.

### 3.3 Creating a PPA delimitation

To create a PPA delimitation based on the division of the original scan file described in the previous chapter, it is just necessary to perform the import of the new Shapefile containing the features one wants to make use of the functionality, present in various GIS systems, known as Buffer. In the case of Quantum GIS, located following menu "Vector / GIS Tools / Buffer (s)."

The buffer is used when one wants to create delimitation vector of distance, user editable, also on a vector data. The image 6 shows the use of the shapefile containing the features of lakes, ponds and reservoirs, since on these a buffer has been generated containing a standard distance of the features

mentioned above.

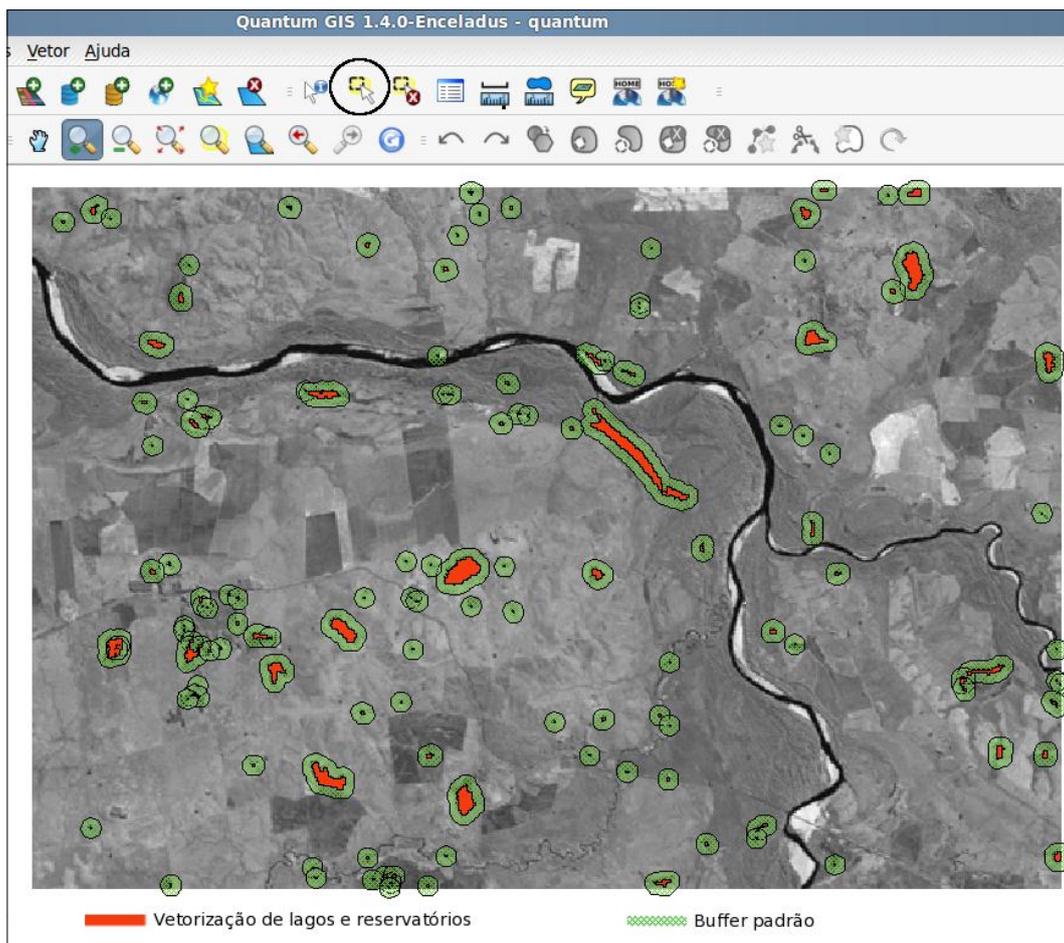


Image 6 - Shape vectoring of lakes and buffer protection.

Likewise it was created the representation of water blades and their buffers protection to lakes, ponds and water reservoirs, in the image 7 the same area is represented containing the display of the watercourse shapefile, and the buffer protection was generated of the same. It is noted that the display of features separately is merely for didactic purpose, not preventing the manipulation of both files at the same time.

#### 4. BUFFER VARIABLE

It is known that, as one of the objectives of this work, we had the proposal to implement a buffer variable to the width of waterways, so we created an image with the specific purpose to serve as a base of base of tests for the development of the second program, this to generate the buffer variable. Besides having as characteristic width variation, a fork was designed in the course of the area of interest to analyze the program's behavior in such situations.

The calculations used were based on the spatial resolution of the image, and specifications for delimitation regarding to the water body of Law No. 12651 of May 25, 2012. It should be noted both the spatial resolution and the distances to be adopted by the buffer, are variable attributes and they can be modified according to user needs.

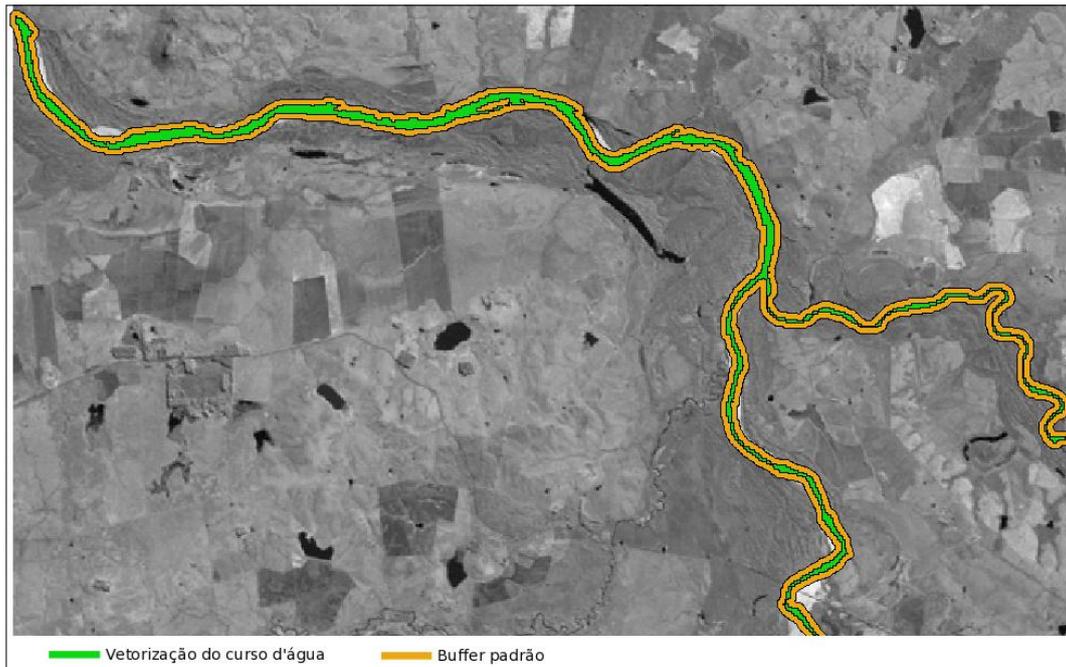


Image 7 - Shape vectoring and stream buffer protection.

As standard distances adopted for the development of the buffer variable, follows specifications drawn from the Article 4 of the law cited in the previous paragraph (BRAZIL, 2012).

- thirty (30) meters: watercourses up to ten (10) meters wide;
- fifty (50) meters: watercourses from ten (10) to fifty (50) meters wide;
- one hundred (100) meters: watercourses from fifty (50) to 200 (two hundred) meters wide;
- 200 (two hundred) meters: water courses from 200 (two hundred) to 600 (six hundred) meters wide;
- five hundred (500) meters: watercourses over 600 (six hundred) meters;

The result of the processing to obtain buffer variable on satellite images can be viewed in the image 8, in which the watercourse is showed and the buffer has its features adapted to the width identified in this course.

## 5. CONCLUSIONS

At the conclusion of this work, we observe that the prototype of vectorization of areas of interest had proven its efficiency. The tests we performed showed a significant gain with respect to performance and accuracy of the product generated from the developed prototype.

We expected the prototype developed during this work inspires and even gives basis for a continuation in the "walk" towards innovation and automation in the area of GIS, both nationally and internationally.

We believe that only with emphasis on research, and by analysis and valuation of critical look at the techniques used in solving problems, it becomes possible the constant human and technological improvement.

From the viabilities expressed in this work, looking forward to new horizons becomes possible. The

results achieved pointed to the possibility to implement future works with the aim of making the features presented until the moment software products, whether in a standalone tool or as modules (parts) from existing programs.

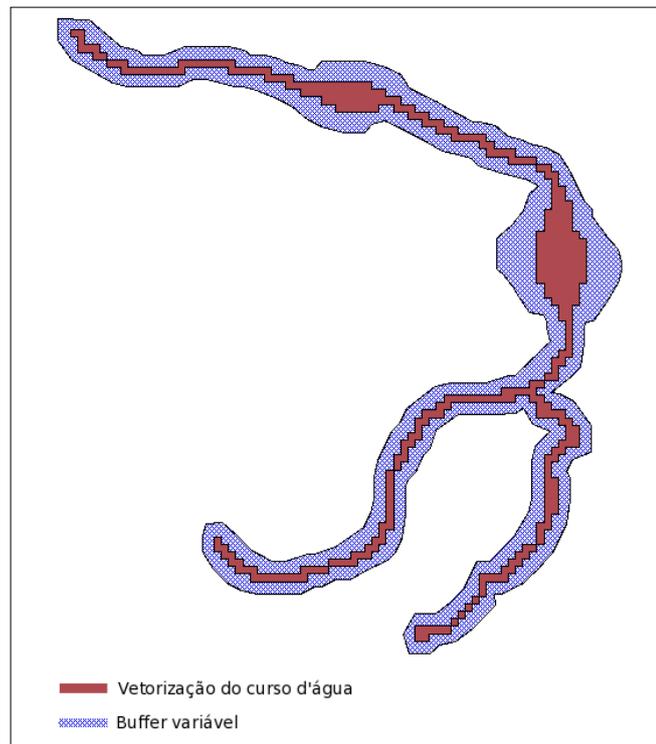


Image 8 – Buffer variable on stream.

## REFERENCES

- BRASIL. **Lei n. 12.651, de 25 de maio de 2012.** Dispõe sobre a proteção da vegetação nativa. **Casa Civil.** Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/\\_Ato2011-2014/2012/Lei/L12651.htm](http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2012/Lei/L12651.htm)>. Acesso em: 11 fev. 2012.
- BORGES, Luiz Eduardo. **Python para Desenvolvedores.** Rio de Janeiro: Edição do Autor, 2010.
- INPE. **Catálogo de Imagens do Instituto Nacional de Pesquisa Espaciais.** Disponível em: <http://www.dgi.inpe.br/CDSR/>. Acesso em: 15 mar. 2011.
- LUNDH, Fredrik. ELLIS, Matthew. **Python Imaging Library Overview.** 2002. PIL 1.1.3 Manual. Disponível em: <http://www.pythonware.com/media/data/pil-handbook.pdf>. Acesso em: 06 mai. 2012.
- MAUBORGNE, R. [Entrevista disponibilizada em 31 de agosto de 2010, a Internet]. Disponível em: <<http://exame.abril.com.br/negocios/gestao/noticias/nem-toda-inovacao-boa-diz-renee-mauborgne>>

## CONTACTS

Msc. Program Graduate Geomatics Éric Tadiello Beltrão  
Universidade Federal de Santa Maria, Av. Roraima, 1000, CEP 97105-900, Santa Maria (RS)  
Santa Maria  
BRAZIL  
Email: ebeltrao@svs.iffarroupilha.edu.br

Ph.D. Forestry Rudiney Soares Pereira  
Universidade Federal de Santa Maria, Av. Roraima, 1000, CEP 97105-900, Santa Maria (RS)  
Santa Maria  
BRAZIL  
Email: rudiney@smail.ufsm.br