University of Bordeaux

Internship Report

Master of Software Engineering (2013 - 2015)

Design and programming of automatic classification methods applied to biological images

---

Student:
LE Van Linh

Supervisor:
Prof. Marie BEURTON-AIMAR

July 24, 2015

Acknowledgements

First of all, I would like to express my deepest gratitude to my supervisor, Prof. Marie BEURTON-AIMAR for her agreement, guide and support during the planning and development of my internship. I would like to thank the staffs in LaBRI, who supported for technique and gave me a professional work environment.

I would also like to thank all professors in University of Bordeaux and PUF-HCM, who imparted a lot of knowledge to learning and researching. I would also like to thank the dean of IT-DLU, who allowed me joined in this course. Finally, I would like to thank my family and colleague for their support and encouragement through my study.

Abstract

Image processing is a field that has many application in life. It can be from the usual application to the application in medicine or cosmology. To obtain the best result, all most of applications must follow two processes: Firstly, we involve primitive operations such as reduce noise, contrast enhancement or image sharpening. Secondly, we can apply the segmentation, description the objects to a form suitable for application process and classification of individual object.

The goal of project is built a program with full functions about processing base on the biological images. During my internship at LaBRI, my tasks are developing the algorithm to preprocessing image and programming of automatic classification methods applied to biological images. After finsihed, I integrated it into the IMP tool, which was developed by NGUYEN Hoang Thao. Besides, we also debug the previous code and write the documentation for the next development.

Contents

1 Introduction **4**
- 1.1 Pôle Universitaire Français . . . 4
  - 1.1.1 PUF-Ha Noi . . . 4
  - 1.1.2 PUF-HCM . . . 4
- 1.2 Laboratoire Bordelais de Recherche en Informatique . . . 5
- 1.3 The Internship . . . 5
  - 1.3.1 Objectives . . . 5
  - 1.3.2 Overview about my task . . . 5
  - 1.3.3 Origanization of the document . . . 5

2 Background **6**
- 2.1 Overview about image processing . . . 6
- 2.2 Smoothing filters . . . 6
  - 2.2.1 Linear filter . . . 6
  - 2.2.2 Order-Statistics filter . . . 6
- 2.3 Image transformation . . . 7
- 2.4 Histogram . . . 7
- 2.5 Segmentation . . . 7
- 2.6 Color processing . . . 7

3 Preprocessing image **8**
- 3.1 Problem . . . 8
- 3.2 Analysis . . . 8
  - 3.2.1 Finding the limit points . . . 9
  - 3.2.2 Replacing the grid . . . 9
- 3.3 Algorithm . . . 9

4 Classification methods **10**
- 4.1 Detect the structure . . . 10
- 4.2 Pairwise geometric histogram . . . 10
- 4.3 Probabilistic Hough Transform . . . 10

5 Implementation **11**
- 5.1 Software architecture . . . 11
- 5.2 Image preprocessing . . . 11
- 5.3 Automatic classification . . . 11
- 5.4 Result . . . 11

6 Conclusion **12**

List of Figures

3.1 The input images with yellow grid 8

Chapter 1

Introduction

1.1 Pôle Universitaire Français

The Pôle Universitaire Français (PUF) was created by the intergovernmental agreement of VietNam and France in October 2004. With ambition is building a linking program between the universities in VietNam and the advanced programs of universities in France. There are two PUF's center in VietNam: Pôle Universitaire Français de l'Universite Nationalé du Vietnam - Ha Noi located in Ha Noi capital (PUF-Ha Noi) and Pôle Universitaire Français de l'Universite Nationalé du Vietnam - Ho Chi Minh Ville located in Ho Chi Minh city (PUF-HCM).

1.1.1 PUF-Ha Noi

PUF-Ha Noi is regarded as a nursery for the linking program, it support on administrative procedure and logistics for the early year of program. About administration, PUF-HN directly under Institut Francophone International (IFI), which was created by VietNam National University at HaNoi in 2012.

1.1.2 PUF-HCM

PUF-HCM[1] is a department of VietNam National Univeristy at Ho Chi Minh city. From the first year of operations, PUF-HCM launched the quality training programs from France in VietNam. With target, bring the programs which designed and evaluated by the international standards for Vietnamese student. PUF-HCM always strive in our training work.
So far, PUF-HCM have five linking programs with the universities in France, and the programs are organized into the subjects: Commerce, Economic, Management and Informatics. In detail:

- Bachelor and Master of Economics : linking program with University of Toulouse 1 Captiole
- Bachelor and Master of Informatics: linking program with University of Bordeaux and University of Paris 6.

The courses in PUF-HCM are provided in French, English and Vietnamese by both Vietnamese and French professors. The highlight of the programs are inspection and diploma was done by the French universities.

[1] http;//pufhcm.edu.vn

1.2 Laboratoire Bordelais de Recherche en Informatique

The Laboratoire Bordelais de Recherche en Informatique (LaBRI)[2] is a research unit associated with the CNRS (URM 5800), the University of Bordeaux and the Bordeaux INP. Since 2002, it has been the partner of Inria. It has significantly increased in staff numbers over recent years. In March 2015, it had a total of 320 members including 113 teaching/research staff (University of Bordeaux and Bordeaux INP), 37 research staff (CNRS and Inria), 22 administrative and technical (University of Bordeaux, Bordeaux INP, CNRS and Inria) and more than 140 doctoral students and post-docs. The LaBRI's missions are: research (pure and applied), technology application and transfer and training.
Today the members of the laboratory are grouped in six teams, each one combining basic research, applied research and technology transfer:

- Combinatorics and Algorithmic
- Image and Sound
- Formal Methods
- Models and Algorithms for Bio-informatics and Data Visualisation
- Programmiing, Networks and Systems
- Supports and Algorithms for High Performance Numerical Applications

1.3 The Internship

The internship is considered a duration to apply the knowledge to the real enviroment. It shows the ability synthesis, evaluation and self-research of student. Besides, the student can be study the experience from the real working. My internship is done under the guidance of Prof. Marie BEURTON-AIMAR in a period of six months at LaBRI laboratory.

1.3.1 Objectives

1.3.2 Overview about my task

1.3.3 Origanization of the document

The all report mainly have five chapters. In the chapter 1, this is the short introduction about my university, mainly information about the lab, where I do the internship and the objectives of my internship. In chapter 2, we talk about the necessary preliminaries in image processing field. In the chapter 3, I propose the algorithm to preprocessing image, with the aim is decrease the noise in the input and increase the effective of the classification methods. In the chapter 4, I mention to the classification process. Finally, I present about the implementation of the preprocessing image algorithm and classification methods.

[2]http://www.labri.fr

Chapter 2

Background

2.1 Overview about image processing

We have a lot of programs what used to edit the photos (e.g. photoshop, gimp, paint,...). By apply some technique, we can effectively some property to change the image such as: scaling image, blurring image, rotating image,.... We also know that, an image is a set of pixels. Each pixel have a value that present for the color at this location, and its location was indicated by coordinates in two-dimension.. When combine the value of all pixels, we have the image as we can see in the real word. The changing on image really changing the value on each pixel in image. Behind the techniques in the programs are mathematical operations and the field using mathematical operation on an input image, called *image processing*. The output of image processing may be either an image or a set of characteristics related to the image. And most of image processing technique are performed on two-dimensional image.

2.2 Smoothing filters

Smoothing filters are used for blurring and noise reduction. This technique is used in preprocessing steps, such as remove some small object unexpected from input image, or bridging of small gaps in lines. Noise reduction can be done by blurring with a linear filter or order-statistics filter.

2.2.1 Linear filter

The idea behind this filter is replacing the value of every pixel in the image by the average of the gray levels in the neighborhood defined by the filter mask. By this work, this filter sometime are called averaging filter. The result of this process is an image with reduced the sharp edges in gray level, it also reduce the noise because the noise is typically and random in the image. The mask is a matrix useful for blurring, sharpening, or edge-detection,.... The output image is accomplished by convoluting between a mask and an image.

2.2.2 Order-Statistics filter

By ordering the pixels in the image and then replacing the value of the center pixel with the value determined by the ranking result. This is the idea of the median filter is the best example used this technique.

2.3 Image transformation

2.4 Histogram

2.5 Segmentation

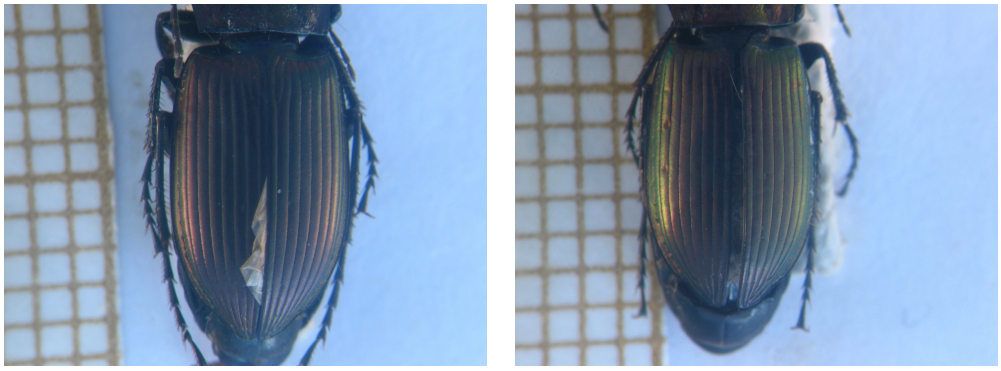
2.6 Color processing

Chapter 3

Preprocessing image

3.1 Problem

The propriety of an algorithm or a program often based on a good input set. To obtain the good result when applying the automatic classification methods. In this chapter, we suggest the algorithm preprocessing image. With the input images contains the parts of insect and an unexpected object, specifically yellow grid (figure 3.1), we need remove the yellow grid to have only insect and just keep the insect.



(a) The yellow gird on the left of insect (b) The insect overlap the yellow grid

Figure 3.1: The input images with yellow grid

3.2 Analysis

Each input image contains the two objects: the part of insect and the yellow grid (called grid). Addition, the grid always stayed in the left of image, and the insect can either overlap the grid or not. About the color, we can see three main groups color: the background color, the yellow color of grid and the color of insect. The image is presented in BGR model. So, the color at each pixel must be combine among three values (blue, green, red). If we process the image in BGR model, the algorithm may be complex. While, the HSV model just has a channel to present the value of color and each color has a clear range. We can apply this property for detecting and removing the gird.
The analysis system is constructed from two main stages: finding the limit points of grid, replacing the yellow point from the begin to the limit point.

3.2.1 Finding the limit points

Browsing all of pixels to checking its color and replacing it if the color of pixel is yellow, we must process on all image. If we do that, it will be waste time. To decreasing the browsing time, in this step we find the limit points. These are the points which located on the right of grid and gird closest.
Finding the limit points will solve the above problem. Instance of checking on all pixel, we just check the pixels stay on the left of limit points. As we know, the width of grid usually less than a two-thirds of width of image. So, to reduce the time to finding the limit point, we also check from the begin of image to two-thirds of image. The result of this step is the limit points, these used for limiting the length when we check the pixels on yellow grid.

3.2.2 Replacing the grid

After having the limit points. By processing on all rows of image. At each row, we replace the pixels which have the color value stay in the range of yellow by another value. But the grid is not only created by the yellow point, it contains more the pixel have the value stay in the same range with background. But the brightness of these pixels is less than the background. So, we needs to replace it obtained the good image. In each row, this work repeated until meeting the limit points or a "special point". It is a point stayed on the insect.

3.3 Algorithm

The algorithm to remove the yellow grid

Data: The image contains the insect and the yellow grid
Result: The image without the yellow grid

1. Converting the input image to HSV model
2. Splitting the image (in HSV) to get the individual channel
3. Finding the limit points
4. Choosing the replace point (the point that its value used for replacing the yellow point)
5. Finding and replacing the yellow points and the "miss brightness" point.

Chapter 4

Classification methods

4.1 Detect the structure

4.2 Pairwise geometric histogram

4.3 Probabilistic Hough Transform

Chapter 5

Implementation

5.1 Software architecture

about architecture of IMP software....

5.2 Image preprocessing

about clear the yellow grid...

5.3 Automatic classification

about methods

5.4 Result

result...

Chapter 6

Conclusion

about conclusion