User requirements

Etienne Caillaud, Thomas Le Bris, Ibrahima Gueye, Gaetan Adier

April 27, 2015

Contents

1 Context

1.1 Context and environment

The project will take place in the XLIM-SIC laboratory of Poitiers University.

It will be about researching for new feature matching and indexing solutions by using new attributes including color and texture features. The final aim of the project will be the participation to the life CLEF contest 2015 which is a competition on the indexing of nature images.

1.2 Aim of project

The project is about identifying and indexing images. The aim is to obtain a new system for image indexing and recognition which is able to work on nature images (that is a huge problem for the standard descriptors which do not include texture or color features). We will have to develop a software program to test the effectiveness of many descriptors. Using this software program, we will be able to compare our results, with the new texture and color descriptors, to obtain the best results in the CLEF challenge.

This project is completely relevant to our training because it includes many subject as image processing and we will have to use many technologies (as Python or C++) to implement it.

There are many educational topics related to this project. We will have to work on project management, do some teamwork and use many image processing skills that we did not know before. We will have to report to our customer and it will be a good exercise for our future professional life.

1.3 Scope of work

We will have to design a software program which, by using the CLEF image learning Database, will be able to index the whole Database according to the different features which are specified in the CLEF rules.

The aim of this project is to adapt the last up to date designed texture and color attributes to the current image classification and then to measure how they perform with photographies taken from nature in order to test their competitiveness through the CLEF challenge.

The XLIM-SIC laboratory is currently working on the design of new image descriptors by integrating texture and color attributes more effectively. This project is a way for the researchers to see how they will behave in a real situation. Furthermore, the start of this project corresponds to the CLEF 2015 contest which is coincidentally about image recognition and automatic indexing of images taken from nature. If the results from this project are satisfying enough, they will be submitted to this international contest.

1.4 Stakeholders

- Customer: Thierry Urruty.

- People involved in the project: Thomas Le Bris, Gaetan Adier, Ibrahima Gueye, Etienne Caillaud, David Helbert, Noel Richard.

- End users: Various image database structures (Flicker, Google, CIBDI Angouleme, Ainden...).

1.5   State of the art

The CLEF contest is a competition between many international laboratories. This competition began in 2010 so we can see what the other laboratories are doing nowadays.

We can see that all the competitors use SIFT or SURF descriptors. In the 2014 contest IBM was the laboratory which had the best results[1]. After analysing what they do, we have concluded that they used all existing descriptors and mix them randomly.

After analysing the result of the 2014 contest we have agreed to use FINKI as a reference [2]. FINKI uses segmentation and after then chooses between two descriptors (multiscale triangular shape or opponent SIFT).

2   Constraints

2.1   Deadlines

The finished product must be submitted by the 9th June 2015 including all necessary documentations.

The presentation will take place on the 15th June 2015.

SCRUM meetings will be scheduled every morning under the lead of Mr Chatellier.

Weekly meetings with tutors will be scheduled according to availabilities and needs

3   Description of the final product

A software program able to calculate a texture type descriptor suitable to nature images and which confirms the performances of this attribute via the CLEF contest results.

Locally, we will have to create one directory for each descriptor which will contain one CSV file in which we will find the descriptors. One text file will be created too and will contain all the information about the image. To facilitate the indexing, one single text file will be created for the whole database and will contain all the information (image → descriptor 1 / descriptor 2 /) as an table of contents. Descriptors must be interchangeable in the software.

When this software program is finished to compute, we will have to compare our results with those of the other laboratories which have participated in the CLEF challenge. We will use the CLEF challenge metrics which are specified in the challenge rules. An abstract of these comparisons will be provided with the software program and the final technical report.

4 Organization and project management

4.1 Schedule

List of tasks:

- Writing the state of the art report

- Learning the use of HULK calculator

- Preparation of the database

    Constitution of database structure

    Calculation of SIFT descriptors

- Programming

    Programming C_2O descriptor calculation

    Programming the distance calculation function

    Writing of the technical documentation

    Calculation of C_2O descriptors

    Compare result using CLEF metrics

- Writing of final report

- Writing of oral presentation

cf. appendix A.

4.2 Place

The project will take place in the project classrooms (ON03/01/17) in the SP2MI building belonging to the University of Poitiers.

5 Necessary resources for development

List of necessary resources for development.

Connection to computing server of the university (HULK).

Computers with Python edition tools are the only necessity (can be found on the ones from the university).

The project participants will have to use their own computers, a place with electrical outlets and Internet access will be sufficient.

References

[1] Pierre Bonnet Souheil Selmi Jean-Franois Molino Daniel Barthelemy Herve Goeau, Alexis Joly and Nozha Boujemaa. Lifeclef plant identification task 2014. In *Information Access Evaluation*, 5th, pages 598 – 615. CLEF, Springer, September 2014.

[2] Rahil Garnavi Xi Liang Qiang Chen, Mani Abedini. Ibm research australia at lifeclef2014: Plant identification task. In *Information Access Evaluation*, 5th, pages 693 – 704. CLEF, Springer, September 2014.
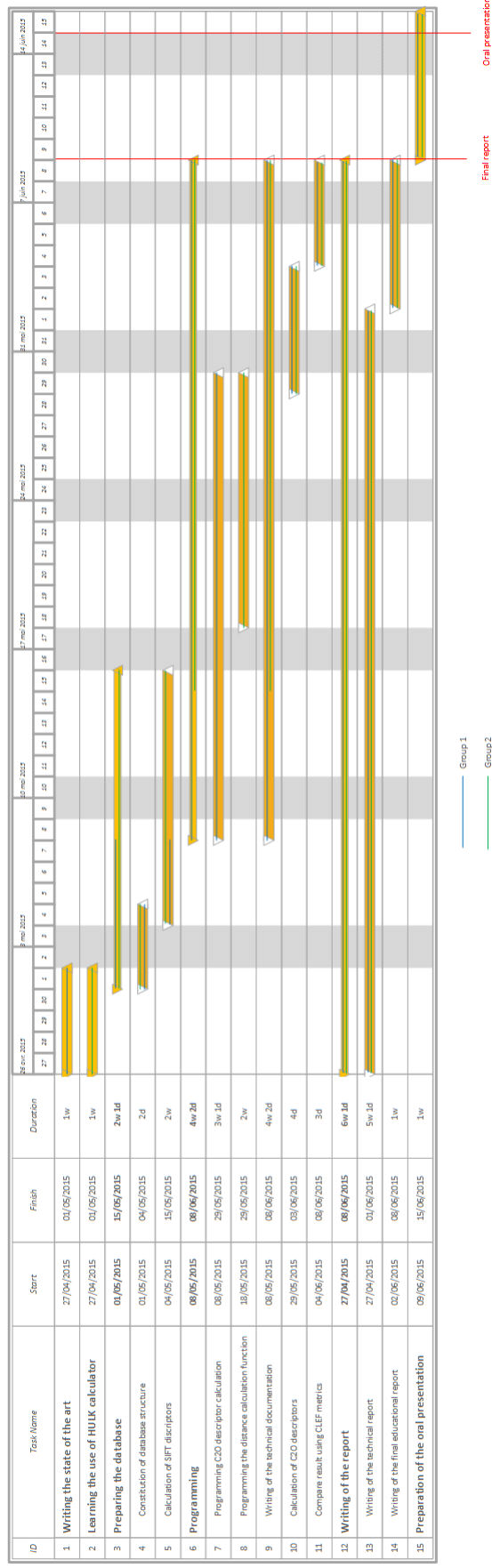
A   Gantt chart

ID	Task Name	Start	Finish	Duration
1	Writing the state of the art	27/04/2015	01/05/2015	1w
2	Learning the use of HULK calculator	27/04/2015	01/05/2015	1w
3	Preparing the database	01/05/2015	15/05/2015	2w 1d
4	Constitution of database structure	01/05/2015	04/05/2015	2d
5	Calculation of SIFT discriptors	04/05/2015	15/05/2015	2w
6	Programming	08/05/2015	08/06/2015	4w 2d
7	Programming C2D descriptor calculation	08/05/2015	29/05/2015	3w 1d
8	Programming the distance calculation function	18/05/2015	29/05/2015	2w
9	Writing of the technical documentation	08/05/2015	08/06/2015	4w 2d
10	Calculation of C2D descriptors	29/05/2015	03/06/2015	4d
11	Compare result using CLEF metrics	04/06/2015	08/06/2015	3d
12	Writing of the report	27/04/2015	08/06/2015	6w 1d
13	Writing of the technical report	27/04/2015	01/06/2015	5w 1d
14	Writing of the final educational report	02/06/2015	08/06/2015	1w
15	Preparation of the oral presentation	09/06/2015	15/06/2015	1w

Figure 1: Gantt chart