

Constitution of the ImagEval Corpus : An end user-oriented approach

Coralie Picault – coralie.picault@neuf.fr
(Laboratoire Paragraphe – Université Paris 8)
TRIBVN

Minutes of the speech delivered on December, 11 during ImagEVAL conference.

TABLE OF CONTENTS

INTRODUCTION (p.1)

1. What Part for an information scientist and image searcher in the project ? (p.1-2)

2. Preliminary step: building up a 158 000 pictures database (p.2-5)

- 2.1. Gathering heterogeneous resources in preparation for making one single database
- 2.2. Holistic understanding of the database contents

3. Building up the 4 learning databases: first comparison between computed approach and human approach (1st stage of the protocol of building up of the corpus) (p.5-6)

- 3.1. Choice of the selection criteria of the images for the tasks: adjustment between the needs defined by the scientific Committee and user's waitings
- 3.2. Building up the 4 training databases

4. Launching of the run tests and checking of the first criteria (2nd stage of the protocol) (p.6-7)

- 4.1. Building up the run test databases: a team work
- 4.2. Results of the run test: 2nd adjustment between “computed” approach and “user” approach

5. Building up the 5 official databases: tasks one by one (p.7-10)

- 5.1. Modified images recognition
- 5.2. Mixed research images/text
- 5.3. Text area detection
- 5.4. Object detection
- 5.5. Semantics extraction

CONCLUSION: towards ImagEVAL 2 ? Critical approach of the process (p.10-11)

The scientific consortium of the ImagEVAL project has as an ambition to conduct a campaign of evaluation of software technologies of automatic treatment of the images by confronting them with the defined needs of professional people in the image's area. Also, to match the goal of the consortium with the pictures' holders, we carried a glance which one will describe as “user” at the time of the constitution of the corpora of images. This is why we will speak throughout the project about end user oriented approach. We will see how we deal with the needs of any involved people that is scientists, holders, users. We will show how each base have vocation to meet these needs.

First of all, we will see what is the role of the information scientist in this project. Then we will be interested in a chronological way in the process of building up the corpora: the preliminary step with the preparation of a 158.000 pictures database; the constitution of training databases after having determined for the first time the selection criteria of the images; the constitution of run test databases and the analysis of the results in order to check the selection criteria of the images and finally the constitution of the official bases, task by task. The set of procedures explained below mainly relates to the tasks of recognition of modified images, of object and text detection and extraction of semantics; the task of research for combined image/text was dedicated to a separate procedure and will be explained later at the end of the meeting.

1. WHAT PART FOR INFORMATION SCIENTIST AND IMAGE SEARCHER IN THE PROJECT ?

a) A specific outlook

The presence of an information researcher having worked in agency brings a different look from that one of computing researchers.

He is useful for:

- building up of the corpus since he has specific skills for this. Thus, to build up a single database starting from heterogeneous databases, **iconographic and information competences** were necessary from the collect of the database among holders to the build up of the new database.

- playing the **role of intermediate** between researchers and end users. A user is an information scientist or an iconographer who search pictures on image databank. In this project, we find again the role of intermediate between researchers and image holders.

b) Acting in concert with researchers

Let's underline that objectives and interests converged towards the will to give functioning tools to the end users. Thanks to this we had a **true partnership** between involved people in the project.

c) Identical working methods

Building up a corpus is **identical** of what has to do an image searcher when he has to look for and index images. An image searcher can rely on specific practices and specific turn in mind to accomplish that work. My role as a specialist was to build up the corpus by combining my knowledge of user oriented images processing with computed image processing.

2. PRELIMINARY STEP : BUILDING UP A 158 000 PICTURES DATABASE

ImageEVAL has as an ambition to have a large database with diversified content that must be a reference database. So we create a database from which we extract several corpus.

2.1. Gathering heterogeneous resources in preparation for making one single database

We first collect images and texts lent by involved image holders:

- Renault
- Photo agency of the Réunion des musées nationaux (RMN)
- Niepce museum
- Archives of Nantes of the ministry of foreign affairs
- Institute of geography

-Patrimoine photographique

-Archiv'images

-Hachette Filipacchi Médias (HFM) group with Gamma, Keystone, HoaQui, Jacana, Explorer and Paris-match funds.

2.1.1. Collection process

a) Image collection

8%¹ of the database was available as of April 2005.

80%² of the database was conveyed during May-June 2005.

The project was already advanced when the 12%³ remainders were collected in September and November 2005.

Finally, Kodak, Editing and Lartigue association did not take part in the project.

b) Data text collection

Data text, whatever they are, title of the photography, name of the photographer, legends, indexings, were collected simultaneously with the pictures, in different way depending on the funds. This collection was not possible for every funds.

For example, for Renault or RMN there was no data text. For some funds, HoaQui for instance, data text were arranged in an excel spreadsheet.

For some others, as it is the case for IPTC⁴ format, data text was embedded in the picture, extractible with specific tools.

c) Remarks and Difficulties

>Data collection has been done in two **different ways**: either data holders send the pictures to us or we had to take it on site which implied extra data handling.

¹ The Institute of geography, Archives of Nantes of the ministry of foreign affairs and Renault funds were quickly available.

²The various funds of Niepce museum, the funds of Patrimoine Photographique and of Archiv'images and the various funds of HFM group with Gamma, Keystone, HoaQui, Jacana, Explorer were conveyed afterwards.

³The funds of RMN (collected in September 2005) and Paris Match of HFM group (collected in November 2005) were the last.

⁴ The IPTC format (International Press telecommunication committee) is a file format used in press area in order to share informations.

This has been time consuming because of many reasons:

- the number of images to export (about 100 000 pictures).

- various information systems. For some, for which it was necessary to learn capabilities and constraints for there query software, as it was the case with the RMN. For Gamma pictures, it was necessary first to export the pictures then, via an other software, we had to extract the data text contained in cards IPTC.

>For some fund holders, the **legal questions** came into play, and criteria from the start when we begin to sort the pictures. Gamma for instance has specific agreement with photographs (same problem with HoaQui).

>Late collect of Paris-Match and the RMN funds was annoying since we needed this fund to feed correctly the database.

2.1.2. Creation of a single database

In order to create an environment approaching the reality as much as possible, we undertook to create one single homogeneous database. This database is accessible through a tool which main functionality is to browse the database by visualizing pictures and associated texts. To reach this, data text had to be gathered in a Access database, when "zoom" and "mosaic" pictures had to be in directories.

This single database aimed to be used for the build up of the corpus since we could query it according to the task.

2.1.3. Processing and homogenisation of the picture and text data

Criteria have been given and protocols have been set up to gather funds in one single homogenous database.

a) Criteria for image processing

The images we collected, did not have the same definition, as standardize was a must, resizing was applied to get a database with pictures with only two formats, zoom and mosaic. This operation was done only for the reduction of the definition, not when the picture definition was lower than requested.

b) Criteria for text processing

As mentioned previously, information available was not identical everywhere: thus, for the funds Explorer, we had various fields for the keywords (keywords known as primary; localisation; people; names; secondary keywords; technical information) while for the funds Gamma, we had a single field keywords.

For Explorer, the keywords were in English whereas at Gamma, they were in French. This heterogeneity in the data is found at various levels (not only on the keywords), this is why we completed a work getting homogenous data texts corresponding to the images.

We thus determined the most relevant fields for:

- the project requirements (for example, for the task of extraction of attributes, the keywords could be used to make queries for each attribute)
- software constraints (free fields for the data).

We then gathered everything in a database Access text.

In short, many data were worked on from the number of the image to the keywords mainly with Excel.

c) Remarks / difficulties

>The work of homogenisation to gain a multi purpose database has been finished in June 2005, (unless Paris Match fund and RMN fund since we got them later).

>This have been a **long and hard work** that needed to be done with care. For example renaming the pictures: each fund has been renamed with a set of number assigned to it. In example the funds Patrimoine Photographique has been renamed with numbers from 00650000 to 00659999. This mechanism enabled us to maintain the characteristic of each funds and locate them, without need of the documentary data attached to them.

>The software used for the database could not deal with the textual information since this field was limited to 256 characters, so we had to find solutions to ensure the coherency and the homogeneity of the textual data.

>We soon noticed that this images databank could not fully be useful to us for the constitution of the corpora using queries. We carried out a methodological test and the result confirms what many professionals and researchers underline for years, namely **impossibility of amalgamating keywords resulting from several methods of indexing**. In this project, the indexing was different for each fund. Although the keywords were located in a single field of the database, which made

queries easier, the indexing was here useless since there was not a single term, given in a documentary language. There were several documentary languages here, in addition in several languages. Once again, we could demonstrate that the melt of the textual could not done without a preliminary work on the documentary languages.

2.2. Taken knowledge of the whole database

2.2.1. Complete browse of database

Before fixing the criteria for each task with the scientific Committee, we decided to browse over the funds in order to have a more precise idea of the contents of the database. Indeed, for an iconographer, to browse over is the primary method to know the contents. The **visual memory** will be essential when the adjustments between the needs for the consortium and those of the users will have to be done. Relying on a good knowledge of part of the funds Hachette, this map of the funds proved to be important since there were already 130 000 images.

2.2.2. Pictures funds map

Get in mind the funds layout.

Kind of pictures	Image holders	Themes of funds
<i>Postcards</i>	Institute ⁵ of geography	West Africa, beginning of the XXth century ; town and population of Tunisia and Morocco, beginning of the XXth century ~4505 images
	Archives of Nantes of the ministry of foreign affairs	town and population of Tunisia ~3095 images
<i>Black and white photographs (archives)</i>	Niepce museum	-Japan fund = town and population of Japon -Lapie fund =

⁵ The Institute of geography run postal cards funds of UNESCO ("Images et mémoires").

		urban aerial photography -Eble fund = portraits -1870 fund= Paris of 1870 ~ 2944 images
	Renault	Cars ; schemas, drawings, plans ; industrial life, 1900-1950 ~ 4724 images
	Patrimoine photographique	Portraits studios, illustration ~999 images
	Keystone (HFM group)	General funds : archives, beginning of the XXth century = news, policy, fashion, sport, monarchies, show-biz (people) ~20100 images
<i>Colour Photographs</i>	Archiv' images	Illustration ~1327 images
	Gamma (HFM group)	General funds : news, show-biz (people), monarchies, fashion, sport, illustration ~80 000 images
	HoaQui (HFM group)	Illustration : travels, landscapes, society, foods ~ 15225 images
	Jacana (HFM group)	Animalist illustration ~1933 images
	Explorer (HFM group)	Illustration like HoaQui ~2046 images

Was only missing the Paris-Match funds which thematic is close to Gamma one (about 4580 images) and the RMN, photographs of pictures, sculptures and art objects (about 15 415 images).

The whole amount will be 158 000, all funds included.

This map must be taken into consideration for the next since it will be primary factor for the constitution of the databases for each task.

This step concluded on June 2005, we had to define criteria to choose the images to put in the training database.

3. BUILDING UP THE 4 LEARNING DATABASES : FIRST COMPARISON BETWEEN COMPUTED APPROACH AND HUMAN APPROACH (1st stage of the protocol)

3.1. Choice of the selection criteria of the images for the tasks: adjustment between the needs defined by the scientific Committee and user's waitings

During the first meeting with Pierre-Alain Moëllic, in charge with the project, the criteria for the choice of the images in the databases were determined while confronting:

- the **“computed” point of view** directly dependent upon technological constraints
- the **“user” point of view** which means the vision of information scientists in their trades with the users. Particularly that refers to the indexer analyzing and indexing the images with the aim of enabling the users to find the images corresponding to their needs and the iconographer to seek the images corresponding to the needs for these same users.

For each task, we thus determined **criteria**, i.e. the visual elements to take into account or on the contrary to avoid on an image. Some criteria were defined for all the tasks:

- variety of the images and the representativeness of the databases
- level of difficulties of the images (concept of feasibility of research).

Then other criteria were specifically defined with each task:

- for the tasks of detection of zones of text or objects, we had to care about the size and the visibility of the zone of text or the object in the image (essential criteria)
- for the task of recognition of modified images, we had to primarily select images with the sharp colours and without presence of text since the transformations used were chromatic or with addition of texts.
- for the task of extraction of attributes, we had to select the images and then classify them in a single semantic class.

The definition of some criteria was prone to discussion, which led us to some **adjustments**. Here are two examples of these adjustments, particularly for the task of detection of the objects:

>The scientific Committee had preset the objects queries. One of the objects preset was the dog. However, we noticed, at the time we browsed the database, that the object “dog” was not enough present in the images in a sufficient and significant number. Thus, in the Jacana fund, fund specialized in animalist photography, only three photographs represented dogs. Moreover these photographs represented too partial and non representative sights to be selected. We thus decided to replace this animal object by one met at first sight more frequently: the cow.

> As a rough guide we had the list of the keywords most often used within the Gamma agency in order to locate the objects that would have been relevant to query. Following the study of this list, we proposed to add certain objects. We agreed on the proposal “black glasses”, useful query primarily for the funds of agencies as well as the funds of illustration.

At this meeting, we also talked about respective quantity of images for each database (training, run test and official test) and for each task⁶.

> Remarks/difficulties -

This phase of choosing the criteria of selection was done without images. We endeavoured each one to define these criteria by calling upon our mental images.

3.2 Building up the 4 training databases

After this meeting, we built up the training databases. We had 2 months to make databases for every 4 tasks.

⁶ Quantity initially provided : -task of recognition of modified images, training database of 60 images ; run database of 250 images ; official database of 2500 images
-task of text detection, training database of 30 images run database of 600 images ; official database of 6000 images
-task of object detection, training database of 500 images ; run database of 3000 images ; official database of 30 000 images
-task of extraction of attributes, training database of 7500 images ; run database of 15 000 images ; official database of 50 000 images

We wondered about **the best way to set up these samples** while respecting the criteria defined previously. They were the first databases to be realized and no method had already been outlined.

We could not proceed by **query**. Also, our first step was **simultaneously to carry out the 4 databases by browsing funds**, one after one.

We realized that this method caused a too important cognitive load: this couldn't be done efficiently that way.

We then go back to a **task by task** process, which need us to mobilize only the criteria for the task we were working on.

A first rather broad selection of the images was carried out from all the funds present.

Of this first selection, we chose the most representative images to constitute the databases of training. Once the latest selection done; we did:

- for the task of recognition of the images modified, the transformations on the images sources⁷
- for the task of text detection, the framing⁸ of the zones of text present in the image, by taking into account of the criteria of fusion defined by the scientific Committee
- cutting of the objects in the image in order to present only the object corresponding for the task of detection of objects.

Training database for the task of recognition of modified images consists of 60 images; the one for the detection of text 30 images. These bases were delivered in August 2005.

Training database for object detection consists of 501 images and the one for the recognition of attributes 5474. These databases were delivered in September 2005.

> Remarks / difficulties

-Although we had only 2 months, we succeeded in constituting relevant samples. They will be **completed** thereafter with the addition of Paris-Match and RMN funds.

-For that kind of project the build up the running test database with tests executed then would have facilitate the training databases constitution.

⁷ These transformations 'll be done for run and official databases.

⁸ The framing of the zone of text 'll be done for run and official databases.

-With these first selections, we noticed that certain selection criteria of the images were still too vague and that we had to choose samples to redefined more precisely these criteria, for instance the size of the text or the size of the object.

-We can retrospectively noticed we proceed with trial and error method, as we already announced, as well as the late awakening of the big amount of work to be realized.

Indeed, for the constitution of these databases, the queries on the data text were not very effective, we had to look for each task the images one by one, so we needed to have more time to complete the work.

4. LAUNCHING OF THE RUN TESTS AND CHECKING OF THE FIRST CRITERIA (2ND stage of the protocol)

Following the sending of the last databases of training, in September 2005, we had already to take into consideration a few problems: the lack of accuracy of some criteria and the requirement of more labour force for the work to be done. Stéphanie Roger joined the project in November 2005 to help to accomplish what need to be done for the launching of the running tests scheduled at the beginning of January 2006.

4.1. Building up the run test databases: a team work

For the building up the databases we had to define the **methodology** to have since there was no previous experience on this.

We splitted up each fund in two, then we work everyone on our side on each task, images one by one. We decided to select the images according to the broadest possible criteria for then compare our choices for each task.

This double check should guarantee the homogeneity of the selection.

The run test databases for the task of recognition of modified images and for the detection of zones of text are made up of 500 images, the ones for the task of object detection is made up of 3001 images and the one for the task of extraction of attributes of 9077 images. These databases were delivered on the beginning of January 2006.

> Remarks / difficulties

-The concept of "**double glance**" is important because it guarantee credibility and reliability during the process of databases building up.

-The choice of the images was more or less easy, depending on :

- the task (the choice of images for the task of recognition of modified images is simpler than for the task of object detection, for instance)

- the function of the images, their informative level. Moreover, we can speak about "simple" images (images with an elementary content) or about "complicated" images (images with a dense content).

The variety of the funds made thus possible to find images answering easily to the criteria defined for each task as well as a whole variety of "complex" images (in particular images of news).

After having knowledge of the images as well as the criteria based upon one hand on a computed orientation and on the other hand on a human orientation, we drew up a typology of the images, particularly for the task of objects detection.

- images which one could describe as "simple", answering as well to the query "user" oriented as "computed" oriented, are the easiest to select because they correspond to the 2 orientations.

The following two types of images are more problematic.

- images answering the "computed" queries but at first sight not very relevant for "human", images selected during the running tests even if they are not images of 1st choice.

- images answering the "human" query, relying on the tacit knowledges of humans or its interpretation symboliquement, but difficult (even impossible) to find for computed query.

- The amount of images initially envisaged for the task of extraction of attributes was lowered, which must push to reconsider the amount of images envisaged for the official tests.

4.2. Results of the run tests: 2nd adjustment between "computed" approach and "user" approach

In April 2006, the results of the run tests were presented at a meeting with the involved laboratories. Starting from examples, we discussed results to ensure a better adequacy

between the possibilities or technological constraints and the user expectations.

We came to the same conclusion: user dimension was to prevail on computed dimension in the selection criteria of the images for the constitution of the databases of the official test. As mentioned previously, the "user" orientation was to take into account the policies of image processing of the various holders of funds including the customer requirements to select the images of the corpora.

This choice agreed with the aims of the project wishing to give a dimension more anchored in the reality of the world of the image databases to the evaluation campaign. Consequently, user dimension prevailed while keeping in mind certain software constraints.

If we chose an image for a query object, in the event of presence of another query object judged non-relevant user point of view, we must keep in mind the software will detect it if it is sufficiently visible. We will have to take it into account despite the least relevance compared to the user's needs.

> Remarks/difficulties

- The **match of "computed" approach and "user" approach** is not always easy to keep because the two approaches are a lot different.

Within the automatic computed framework, the indexing is carried out according to invariant algorithms (unless intentional modification) from one image to another which allow a relatively uniform classification. Within the documentary and iconographic framework, the user has his own glance which calls upon its sensitivity partly, with its culture which is obviously specific to each individual. This glance exists for user but also for information scientists. In this case, remains the **question of subjectivity** although we were two people to do the selections in order to moderate the subjectivity of the other and to confirm choices. Moreover, any image considered to be too ambiguous was removed.

5. BUILDING UP THE 5 OFFICIAL DATABASES, TASKS ONE BY ONE (3rd stage of the protocol)

Following this meeting in April 2006, we had two months to finalize the databases for the official test.

We proceeded as we did for the building up of the run databases. We browsed the whole funds, to choose the relevant images.

The database for the recognition of modified images is made up of 3000 images, that one for combined research text/image is made up of 6000 images, that one for the detection of zone of text of 500 images, that for the object detection of 14.000 images and finally that one for the extraction of attributes of 23.573 images. These databases were delivered in June 2006.

The build up of the databases was done by taking into account of the various criteria defined before with the point upstream, during previous phases of the project. We will be interested now in the criteria selected, task by task.

Two things were to be taken into account for all the tasks:

- the diversity of the images in the corpora: we wish that databases of evaluation resemble as much as possible to usual databases. We thus took care of respecting the representativeness of the various funds.
- the level of difficulty of the images: we were to take care of the level of difficulties of the images.

5.1. Modified images recognition

The choice of the images for the task of recognition of modified images was made according to following criteria's:

- **prevalence of the color images** : we selected mainly color images because the transformations were mainly chromatics. Images with sharp colors were thus recommended and the manually colorized images were to be avoided.
- **themes diversity of the images**: we took images **from all the possible domains** because it was unnecessary that they are too close. Moreover, we were to ensure the representativeness of the database.
- **non-attendance of text** : we were to avoid the images including text because certain transformations were based upon text.
- a quota of **similar images**: to slightly increase the level of difficulty of the task, we selected some visually similar images.

> Remarks / difficulties

To ensure the representativeness of the database contents, we applied quotas according firstly to the types of images (color images; black and white images; postcards; reproductions of works of art); secondly the contents of the images (sport; mode; showbiz; policy; news; illustration)

5.2. Mixed research text/image

> Steps for this task

For this task of combined research text/image, we worked in **close cooperation** with the scientific group working on the project. We first collected a set of possible queries then we selected some of them according to their picturability. For example, we selected famous monuments like the Eiffel Tower. To this set of queries, we associated other more or less close keywords in order to obtain a reasonable amount of Web pages and a sufficient complexity for the task. We then used two traditional search engines on Internet: Google and Alltheweb in order to get a list of URLs to create the official database. To sort out all the pages in order to obtain, for each one among them, a text file and the images files, we used a tool provided by the scientific organizer. Lastly, the relevance of each image was analyzed in order to match the real world

> Criteria for the choice of the images

The choice of the corpus for the task of combined research image text was done according to following criteria's:

- the **quality of the visual representation** of the queries: we retained **queries of an encyclopaedic nature** having a good potential of visual representation. We thus selected categories like:
 - famous sites or monuments: Eiffel Tower, Statue of Freedom, Ayers Rock (Uluru rock), Bored Rock, Great Wall of China, Niagara Falls
 - symbolic entities: European flag, Ethiopian flag, map of Norway
 - animals: ladybird, zebra, Siamese cat, clownfish, bee
 - works of art: Mona Lisa, Guernica, Freedom guiding the people
 - objects (natural or artifacts): poplar, lava flow, lemon, lawyer, ball of tennis, tennis court, teddy bear, screwdriver

- **size of the images**: we filtered the images according to their size, which made possible to reject a significant number of icons in the Web pages which were without interest for the evaluation.
- the **simplicity of structure of the Web pages**: we retained “simple” Web pages, by avoiding pages with frames.

5.3. Text area detection

The choice of the images for the corpus of the task of detection of the zones of text was done according to following criteria's:

- **concept of typographical character** : we agreed on what we call text, typographical character, namely a minimum of two letters or two digits. The characters could thus be **alphanumeric**.

Any types of alphabets (Latin, ...), any languages (French, English, Chinese, Arabic,...) were accepted.

The writing could be **typographical or handwritten, in capital letter or not**.

- the **size and the visibility of the zone of text**: the text was to be easily **readable with the naked eye**.
- the **number of zones of text**: we could retain images with one or more zones of text. On the other hand, we were to avoid the **overlapping** between various zones of text and the zones too complex, especially compared to the size and with the lisibility (for example, a plastic bag on which the brand of the store is written).
- the **absence of text in the image**: we had also to take images without text.

> Remarks / difficulties

- the criteria of fusion determined by the scientific Committee made difficult the preparatory work of the bases for this task. It was complicated to determine when there were several zones of text in the image, if these zones overlapped or not. This work is not obvious with the naked eye since it is mentally necessary to calculate the distance between two zones of text.
- then another difficulty was to adjust our threshold of lisibility to match that one given for the software.
- lastly, amount of images has been re-examined down because of the level of difficulty of the task.

5.4. Object detection

For this task, the funds of the RMN was littlely used. Indeed, we did not use the drawings or paints possibly representing one of the objects.

The selection of the images was carried out by taking account of the following criteria:

- **visibility and size of the object**: the object must be locatable and recognizable with the naked eye.
- **presentation /angles of shots of the object**: the object may be presented under all its profiles: face, side, back. However the shots in "bird eye" or low-angle are avoided.
- **occlusion of the object**: the object can be presented partially: front part of a plane; back part of a car; upper part of the Eiffel tower... It can also be obstructed by other elements: several personalities in front of a car , ...
- **Centring**: the object can be taken according to various points of view while always respecting the criterion of visibility and recognition of the object. It can integrate a general ground just like it can be the object of a close-up.
- the **nature of images representing the object**: the drawings or paintings were not retained except one or two for the US flag. The choice was made among the color photographs, black and white photographs as well as the postcards.
- the **absence of query object on the image**: we were to select images without objects queries to add difficulty.

These criteria are the criteria common to all the objects. In addition to these criteria, others were defined object by object⁹. These criteria vary according to the **univocal** character of the object. According to this, the definition given to the object was more or less accurate. Thus, the criteria of selection for the “car” object are much more numerous than for the object “Eiffel Tower”, for example. For the “car” object, criteria of physical shape of the object (different shapes from Twingo to the 4x4), of date of the object, places of representation (from the assembly line to a simple road while passing by the car show) were defined. It is not the case for the Eiffel tower, univocal object and from which we had excluded the models. The place of representation was the only modifiable criterion.

⁹ To get more détails, the reader is referred to the document *protocole_choix_t0405* available on : <http://www.imageval.org>

The univocal objects needed less work to define the criteria than the objects having more possible images as the tree.

> Remarks / difficulties

- we could not reach the amount of images which had been intended at the meeting in March 2006 in spite of the number of already important documents of the database and the additional images carried out in May 2006 with the Gamma funds. If we had increased the number of objects, the quality of the corpus and its user-oriented main characteristic would have been partially lost. We provided a corpus of 14000 images instead of the 30.000 images intended initially.

- in addition, this task led us to ask many questions about these images and their integration in the corpus. If we refer to the typology evoked previously on the images, we mean relevant user-oriented images. We found according to some elements, some images that could have been relevant. However because of a high level of complexity, we eliminated these images considered to be too complex or ambiguous.

5.5. Semantics extraction

The distribution of the images was carried out by following the next criteria:

- **membership to a category**: each image was classified in a semantic. In the case of ambiguous image (some of the ambiguous images were not excluded), the choice was made by indexers while respecting to the maximum the logic defined at the beginning for each semantics.

- **membership of an image to a precise semantics**: the partitioning of semantics was organized according to a tree structure where on the one hand, two semantic are unique, generic (reproduction of works of art and colorized images) and where on the other hand, two semantic has specific categories (the color photographs and black and white photographs) have the semantic particular ones: inside / outside / day /night; urban scene/natural scene).

Following the meeting in April 2006, we decided that each image should belong to a specific semantics (example: an image cannot be only classified in semantics "color").

For this task, we had also to determine specific criteria to specific semantic¹⁰. For example we determined how to classify the images between semantic "day" and semantics "night" for the tendentious images. Indeed, images being between the two semantics (sunset or sunrise, ...) are prone to subjectivity. We thus defined criteria for these semantic for homogeneity and reliability concern of the corpus. For example, if the sun were still present on the image, we classified the images under semantics "day". In addition, primarily for the urban scenes, we classified the images under semantics "night" as soon as we could see the lights of the city.

> Remarks / difficulties

-we were confronted with the **imbalance amount of images** for each semantic. Indeed, we could largely reach the amount requested for certain semantic, like that of the indoor; on the other hand, for others, we could not get the amount.

We did not reach total amount because we could not create an unbalanced corpus. In addition, the obligatory membership of a precise semantics contributed to the fall of amount for this corpus.

- certain photographs could not be simply classified in the semantic precise ones, in particular because of subjectivity. Also, we eliminated certain images considered to be ambiguous.

CONCLUSION: TOWARDS IMAGEVAL 2 ? CRITICAL APPROACH OF THE PROCESS

The building up of the database then corpora took place from April 2005 to June 2006. We can now give an assessment on the whole processes which were developed in this project.

> **No preliminary set method available** : the working method initiated for ImageVAL was developed day after day except for the method of gathering data which has already been before by Tribvn within another project. We can observe retrospectively that methodological tests were not convincing. We could not count only on the data text of images database to create the corpora.

¹⁰ To get more details, the reader is referred to the document *Protocole_choix_f0405* available on : <http://www.imageval.org>

We can however consider this method as valid while not forget it is not free from problems.

> **Time management problem:** we should undoubtedly have devoted more time at the beginning of the project for tuning the queries with the funds contents. Although the sampling database could be richer, this phase of constitution of the bases of training requires a more time. The queries were selected a priori and we should undoubtedly have chosen them in agreement with the contents of the databases (example: the query “dog” was removed).

> **Reflexion regarding the funds** to be chosen should be done. The majority of the images was images of news, these images are used daily by the users but they are complex images with many iconographic elements. For some tasks, one would need more limited funds, with simpler images. The real world funds are not always usable just as they are for the automatic tools.

> **Problem with the tools:** the provided tools were not the best one for the workload. For some handling, we will have probably to implement more adequate and more powerful tools.

> **Underestimated workload:** the constitution of the corpora required an additional person to finish the job greatly in the deadline.

> **Problem of amount of data:** for some tasks, required amount for each database was too important if we wish to respect the criteria defined to choose the images. We had to reduce them.

> **Proposal for the tasks:** for the task of recognition of the modified images, it would be interesting to insert press articles in which the photographs were published. The photographic organizations must find the name of the photographer and the number of image for the invoicing and the payment of the photographers. The practical interest is thus immediate. We had mentioned it at the beginning of this project but this proposal could not have been adopted. For the task of object detection, it would be interesting to introduce more objects meeting the user's needs.