Evaluating Content Based Image Retrieval Techniques with the One Million Images CLIC TestBed

Pierre-Alain Moëllic, Patrick Hède, Gregory Grefenstette, Christophe Millet

Abstract—Pattern recognition and image recognition methods are commonly developed and tested using testbeds, which contain known responses to a query set. Until now, testbeds available for image analysis and content-based image retrieval (CBIR) have been scarce and small-scale. Here we present the one million images CEA-List Image Collection (CLIC) testbed that we have produced, and report on our use of this testbed to evaluate image analysis merging techniques. This testbed will soon be made publicly available through the EU MUSCLE Network of Excellence.

Keywords—CBIR, CLIC, evaluation, image indexing and retrieval, testbed.

I. INTRODUCTION

PATTERN recognition and image recognition techniques are usually developed and tested using testbeds. A testbed for content-based retrieval of text or images consists of a list of queries, a set of items (documents, images, videos, sound recordings, etc.) and a mapping between the queries and the items. The mapping specifies which items are relevant to which queries. Such testbeds permit the calculation of recall and precision statistics of the recognition techniques used, and thus to evaluate and compare different approaches. Though large-scale testbeds have been created for text, current testbeds for testing image recognition techniques in content-based image retrieval (CBIR) are scarce and small-scale (among the most famous and used: Image databases of University of Columbia [1], Corel database [2], Texture databases like Vistex [3]).

Very large testbeds create a number of problems. Precision and recall of systems can decrease as the discrimination powers of pattern recognition algorithms are pushed to their limits. Processing times are also put to the test. People expect answers from search systems in a matter of seconds. But such problems must be solved as both individual and industrial users are creating larger image collections due to the banalization of electronic imaging. Large-scale image manipulation applications include: online sales of visual content, technological watch, and management of photographs collections (of companies, museums, etc.). The lack of large testbeds that replicate current complexity and size weakens the claims that Content Based Image Retrieval (CBIR) methods can be useful for real-world tasks. Among the different systems, we can quote Qbic[4], Blobworld [5], VisualSeek [6], SIMPLIcity [7] or Ikona [8]

Problems that hinder the constructions of such large-scale testbeds for research is the collection of royalty-free images, end the hand labelling of the large amount of images that are needed for the testbed. Our solution to these problems is the CLIC testbed: CEA-List Image Collection (List: Laboratory of Integration of Systems and Technologies). To create our large-scale testbed, we first hand-labelled a kernel of 15,200 images, assigning them to semantic classes. From this kernel, we then generated one-million distinct, but labelled, images using a variety of general image transformations (geometric, chromatic, etc.) described below. This architecture of kernel-and-variations allows researchers to use the CLIC testbed in many ways and test their systems along several criteria: classical recall and precision, invariance tests (rotations, chromatic distortions, etc.), analyses of processing times, tests in automatic classifications, etc.

The next section presents the steps in the construction and the final characteristics of the CLIC testbed: the global composition, the organization of the 15,200 images kernel, the description of the transformations using for the generation of the million images, the nomenclature and structure of the base and the future evolution of CLIC. In section III, we present the condition of use and different way to use the CLIC base. Section IV describes our initial experimental results over the CLIC testbed with our image indexing and retrieval system *PiRiA* [9]. This is followed by a conclusion.

II. THE CLIC TESTBED

A. Global composition

CLIC is composed of a kernel of 15,200 images and a complete testbed of one million images generated from this kernel.

Fig 1. Two images from CLIC. Left: an original image from the kernel ("Mountain" class); right: a transformed image (negative transformation) in the testbed

B. Composition of the kernel

The kernel is composed of 15,200 images, which were donated by employees of the CEA-List. These images are completely royalty-free (see *III.A*) for research purposes. The images are representative of images taken by common numerical cameras. They represent outdoor or indoor scenes, natural or urban landscapes, objects, as well as synthetic images. Several additional classes included in CLIC represent signs and symbols (flags and roadsigns).
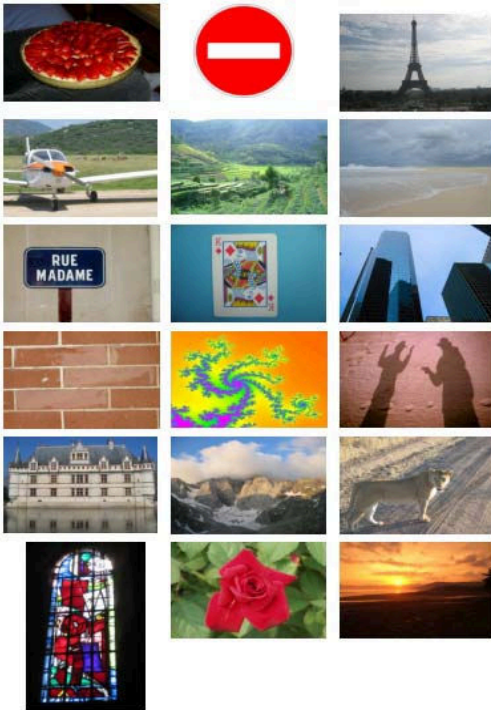

Fig 2. Some images from classes of the kernel of CLIC

The 15,200 kernel images (as are the entire 1 million images) are stored in JPEG format. The original donated images have been resized to 256x384 (384x256) pixels except for one category (roadsigns). The 15,200 images have been manually grouped into 16 major classes, some of which contain subclasses. Here it the list of the major classes:

Food: images of food, and meals.

Architecture: images of architecture, architectural details, castles, churches, Asian temples.

Arts: paintings, sculptures, stained glass, engravings

Botanic: various plants, trees, flowers.

Linguistic: images containing text areas.

Mathematics: fractals.

Music: images of musical instruments

Objects: images representing everyday objects such as coins, scissors, …

Nature&Landscapes: landscapes, valley, hills, deserts, etc.

Society: images with people.

Sports&Games: stadiums, items from games and sports

Symbols: iconic symbols, roadsigns, national flags (real and synthetic images)

Technical: images involving transportation, robotics, computer science.

Textures: rock, sky, grass, wall, sand, etc.

City: buildings, roads, streets, etc.

Zoology: images of animals (mammals, reptiles, bird, fish).

C. Image transformation

Each image in the kernel underwent 49 transformations (see below) to produce 49 new images. Ten of these transformations were applied to two of these new images (a black-and-white version of the original image and a 256-color version of the original image), thus generating 20 additional images. Each original image thus generates 69 additional transformed images. The newly generated images were stored in the same class and subclass as the original image, and therefore inherit the same class label. The difference between a kernel image and a transformed image can be easily recovered in the naming convention used for the new image.

The transformations applied to each original kernel image are the following:

Basic Transformations:
- Entropic thresholding.
- Color histogram equalization
- Linear normalization ([min,max] to [0,255])

Geometric transformations:
- 18 Rotations: from 9° every 10°
- Translations in eight directions, with the norm of each translation randomly computed.
- Horizontal split
- Vertical split
- Transposition
- Projection on a inclined plane

Chromatic transformations:
- Negative
- Black and white (mean of R, G, B values)
- Quantification: 256 colors
- Reduction of the saturation
- XOR operation on a quarter of the image.

Filtering transformations:
- Smooth (low pass)
- Noise (random)
- Gradient (high pass)

Other transformations
- Text incrustation: the word ‹CLIC› located in the centre of the image.
- Border incrustation: random thickness from 10 to 20 pixels.
- Crop: Squared window. Random size and position.
- Mosaic effect on a 4x4 paving of the image.
- Size modification: 64x64 pixels (iconic format), 128x128 pixels, reduction of 25%, increase of 33% (bilinear interpolation).

D. Nomenclature

For the kernel of 15,200 images, the name of the classes and subclasses are in English. The image name is composed of the prefix "clic", underscore, a 3-letter class identifier (only the major class name is used) and a 5-digit number. Thus, the 403rd image of the class "Animals" is named: *clic_ani00403.jpg.*

For the transformed images, an additional 2-digit number corresponding to the applied transformation is affixed. An index file contains the correspondence table between this number and the description of the transformation. For example, the image representing the mosaic effect on the second image of "City" will have: *clic_cit00002_66.jpg*

E. Size of CLIC

The complete CLIC testbed is composed of 1,064,000 images, occupying 50 Gbytes on disk.

F. Evolution of the CLIC tesbed

We plan on producing future version of CLIC in order by increasing the number of images in the kernel, deepening the classification, and implementing additional transformations.

III. Uses

A. Condition of use

CLIC has been built for advancing research in the scientific community. It is composed of images that are royalty-free for research. Research groups can freely use CLIC for publication or for public demonstration. For any publication based on the CLIC database, the name of the testbed (CLIC) and a reference to this current paper must be included.

B. A multi-use testbed

The main objective of the CLIC tesbed is to allow research groups in pattern and image recognition to test their algorithms with a very large testbed composed of wide variety of classified images. With one million images, the CLIC testbed makes it possible to test algorithms against real-world size database. Here, we define six different uses of using the CLIC database.

C. Classical CBIR evaluation with the kernel

The kernel of CLIC can be used for a classical CBIR evaluation (Recall/Precision) using the different classes, or some subpart. The task to perform is: given a photo as a query, find all the relevant photos of the same class (or subclass).

D. Evaluation of the behavior towards the size of base

This classical evaluation can be enhanced by increasing the size of the database and analyzing the quality of answers according to the volume of data. With CLIC, test collection size can vary for 15,200 (or less if we only consider a part of the kernel) to more than 1,000,000 images, allowing evaluation both of the discriminative power of underlying image recognition and of processing time increase with volume of data.

E. Invariance of algorithms according to transformations

Any photo in CLIC can be used as a query with 69 relevant images to be found among the 1 million image database. This task permits the evaluation of algorithm invariance to several kinds of transformations. Such evaluation is important for many commercial applications and more particularly in copyright protection.

F. Automatic classification

Several classes of CLIC have been built to allow evaluation of classification techniques, especially according to attributes concerning the nature and context of the image, for instance:
- Photographs / Clipart
- Indoor / Outdoor
- City / Nature
- Presence / Absence of people

G. Objects, people detection

For some classes, the images represent one or several objects or people. The data can be used for classical object recognition (car, tree, glass, airplane, etc.) and people detection (skin segmentation, face detection, etc.).

H. Detection and extraction of text areas in images (OCR)

The classes "Symbols" and "Linguistic" are composed of images contain text areas. About 400 images, corresponding to different level of complexity, can be used to evaluate techniques of detection and extraction of text in images (OCR).

IV. Some experiments with the CLIC Testbed

We present some examples of uses and initial experimental results with our CBIR systems *PiRiA*. This system considers several indexers dealing with color (HSV global histogram and local HSV histogram with a morphologic region based segmentation), texture (local descriptor histogram) and shape (Fourier) with the possibility of merging the characteristics (for instance Color/Texture or Texture/Shape indexation). Tests with CLIC show that *PiRiA* reach the rate of 580,000 images/second for the research process, that is to say 1.8 second for the million testbed. The indexation process (color/texture) takes 0.15 second/image (image: 256x384 pixels), that is to say 1.85 day to index CLIC.

The present results are for the following evaluation:
- Classical recall/precision on the kernel.

- Invariance (all the base)
- Automatic classification.
- Skin detection

A. Classical recall/precision

We consider 50 images taken in different classes of the kernel and we compute the recall and the precision (considering the 25 first answers).

CLIC	Precision (on the 25 first answers)	Recall
Kernel	0.52	0.29

B. Invariance.

We consider the million testbed. We took 50 images from the kernel as queries. This evaluation only deals with geometric, chromatic and filtering transformations. We consider the 38 first answer and we compute the precision: for each image there are 38 relevant images corresponding to the 38 transformations (Precision = Recall):

$$\text{Precision} = 27.6\ \%$$

Problems come from the nature of the indexer (global indexers on color and texture characteristics). Results are better for geometric transformations than chromatic transformations.

C. Automatic classification

Here, we focus the classification on 3 kinds of attributes: Photo/Clipart, Indoor/Outdoor, and City/Nature.

We built 6 sets of images corresponding to these attributes. "Photo" regroups a 1,000 sample of the kernel, "Clipart" regroups 550 images from category "Symbols", "Indoor" regroups 320 images of categories "Architecture" and "Indoor", "Outdoor" (1,000) regroups categories "City" and "Nature& Landscapes", "Nature" regroups 2600 images from "Nature&Landscapes".

The algorithm uses color and texture characteristics and learning process (Support Vector Machine) with a learning database composed of 1200 images of indoor and 1200 images of outdoor collected on Internet.

Classification	Photo/Clipart	Indoor/Outdoor	City/Nature
% Correct classification	Photo: 98	Indoor: 89.8	City: 94.2
	Clipart: 93	Outdoor: 90.8	Nat.: 92.1

D. Skin (people) detection

We use images from the category "Society" composed of images with people. The algorithm uses 5 conditions applied on R,G,B (normalized) components [10].

Fig 3. Skin detection for two images of the category "People".

V. CONCLUSION

We have presented the CLIC testbed composed of one million images which we feel is a needed resources for the scientific community involved in content based image retrieval and image analysis research. CLIC is composed of royalty-free-for-research images and will be made available to the entire scientific community in computer vision. The generation of the CLIC testbed, by automatically generating images from a classified kernel, makes it possible to create a great number of images, mitigating usual problems in the construction of large testbeds. CLIC has been designed to offer research groups a testbed presenting several possible used to evaluate different kind of image processing algorithms. The different transformations used to generate the million images allow systems to effectively measure the behavior of their systems on real-world size databases, and to prove the invariance of their algorithms to common transformations.

We have also described some initial results using our own image processing system *PiRiA* using this same database, illustrating some of the different uses and interest of our one million images CLIC testbed.

REFERENCES

[1] http://www1.cs.columbia.edu/CAVE/
[2] http://wang.ist.psu.edu/docs/related/
[3] http://vismod.media.mit.edu/vismod/imagery/VisionTexture/vistex.html
[4] M.Flickners, H.Sawhney, W.Niblack, J.Ashley. "Query by image and video content: the Qbic system". *IEEE Computer*, September 1995.
[5] C.Carson, S.Belongie, H.Greenspan, J.Malik. "Blobworld : Image segmentation using expectation-maximization and its application to image querying". February 1999.
[6] J.Smith, S.Chang "Querying by color regions using the visualseek content-based visual query system". *Intelligence Multimedia Information Retrieval AAAI Press*, 1997.
[7] James Z Wang, Jia Li, Gio Wied erhold, ``SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture Libraries,'' *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 9, pp. 947-963, 2001.
[8] N. Boujemaa, J. Fauqueur, M. Ferecatu "IKONA: Interactive Generic and Specific Image Retrieval". *International workshop on Multimedia Content-Based Indexing and Retrieval*, 2001, Rocquencourt, France.
[9] M.Joint, P.A. Moëllic, P. Hède, P. Adam. "PIRIA: A General Tool for Indexing, Search and Retrieval of Multimedia Content". *SPIE.Electronic Imaging*, Vol. 5298, Algorithms and Systems III, Session 3, San Jose 2004.
[10] Chen-Chin Chiang, Wen-Kai Tai, Mau-Tsuen Yang, "A novel method for detecting lips, eyes and faces in real time". *Real-Time Imaging 9*, pp. 277-287, 2003.