Projet ANR- 11-IS02-001

MEX-CULTURE/ Multimedia libraries indexing for the preservation and dissemination of the Mexican Culture

Deliverable

Final report on scalable visual descriptors

Programme Blanc International II- 2011 Edition

A IDENTIFICATION

Project acronym	MEX-CULTURE
Project title	Multimedia libraries indexing for the preservation and dissemination of the Mexican Culture
Coordinator of the French part of the project (company/organization)	Centre d'Etude et de Recherche en Informatique et Communications – Conservatoire National des Arts et Métiers
Coordinator of the Mexican part of the project (company/organization)	Centro de Investigación y Desarrollo de Tecnología Digital – Instituto Politécnico Nacional
Project coordinator (if applicable)	Michel Crucianu : France Mireya Saraí García-Vázquez : México
Project start date Project end date	01/01/2012* 30/04/2016
Competitiveness cluster labels and contacts (cluster, name and e-mail of contact)	Cap Digital Paris-Région Philippe Roy Philippe.Roy@capdigital.com
Project website if applicable	http://mxcprj.hopto.org/

** The Mexican partners are only financed since November 2012.*

Coordinator of this report	
Title, first name, surname	*Prof. Mireya Saraí García-Vázquez*
Telephone	*(+52) 664 3 47 21 00*
E-mail	*freemgarciav@gmail.com*
Date of writing	*10/05/2016*

Redactors :	IPN

B Introduction

Within the project, we focused on content-based search for images and video. For image retrieval, global features especially concerning color are widely used because color is visually very relevant and is insensitive to image translation, scaling and rotation.

The Content Based Image Retrieval CBIR has appeared in 90 years. It represents each image by a set of visual low-level features such as color, texture, shape and movement. These visual features are calculated automatically and then exploited for the system to compare and retrieve images.

Content-based indexing requires the definition of generic content-based descriptors, as well as devising specific detectors for relevant components of the cultural content. Within Mex-Culture, for image and video content, specific detectors were devised for elements of nature, human figures and Mexican architectural structures.

C Type of Mexican nature

Cross-media description of visual content is the process of generating visuals descriptors by exploiting the media streams (images, videos), belonging to a single document. The methods for obtaining visual descriptors are based on the analysis of information from the visual content of the images and videos. This analysis generates features of color, texture, shape and motion, or a combination of these.

To identify the type of Mexican nature in a scene, the **elements of nature** such as different types of vegetation, water, sky and land type are identified. The percentage of items found can determine the type of nature in the scene. We defined a method for detecting "sky", "water", "abundant vegetation" and "sparse vegetation". This requires the description of image patches with statistical moments computed in the color channels [ED1-1], then feeding these descriptors into neural networks [C.1.1]. Fig. 1 shows the process of detection of elements of nature. According to the system of Fig. 1, detection of nature's elements was performed with training of the neural network feedforward type and the neural network with radial base transfer function [C.1.1] to classify four categories of elements of nature in the image content.

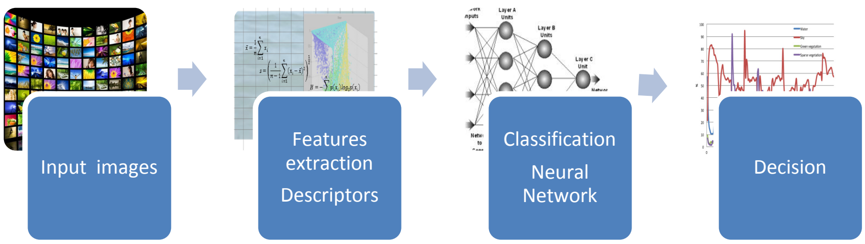


Figure 1: Detection elements of the nature.

Following Fig. 1, the **detection of nature's elements** was performed to classify into four categories the elements of nature in the image content. For each category 150 images were generated, thereby having a training database of 600 video images extracted from the Mex-

Culture database. The methodology and the results are being described in a paper sent to an international journal.

D HUMAN FIGURES

The detection of **human figures** in the images is done by skin detection. Descriptors are extracted from each image with the information of the color components of the 4 models RGB, YIQ, YCbCr and HSV to obtain the information of skin color [D.1.1]. The results show a detection rate of 81%.

E CLASIFICATION OF IMAGES CONTAIN ARCHITECTURAL STRUCTURAL CONTENT

Content Based Video Retrieval Systems (CBVR) select extracted features from video content for selecting, indexing and ranking according to the potential interest to the user.

Our contribution is to provide a method for the classification of images that contain **architectural structural content**. This is performed through the combination of the shape information, viewpoints (vanishing points) [E.1.1] and points of interest. The proposed method is based on the extraction of geometric features containing information about corners and lines detected, the intersection of lines and the ratio of corners and lines. Using the edges map and the Hough transformation, the lines of the image are obtained, on this processing technique we calculate a relation of the minimum line length that will be detected, in relation with the image dimensions. Intersections of the lines found on the image are computed in order to obtain an approximation of the vanishing points [E.1.1]. With this method, we are looking for the relations between lines and corners [E.1.2], using the Canny edge detector and the Hough transform.

E.1 ESTIMATION OF VANISHING POINTS

A set of parallel lines located on the 3D space are represented on the 2D space on an image taken by a camera in a projective space, these lines tend to a common point called vanishing point, this is the point where the lines make an intersection.

However, this does not mean that this intersection of two lines placed on the 2D represent parallelism on the 3D space [E.1.3]. The vanishing points can be used on several applications of computer vision, for example, this approximation to the vanishing points let us a reference to obtain the location of interest points on images in order to implement methods of 3D reconstruction or other techniques of image processing.

With the calculus of the intersection of the lines, a map of intersections is obtained around of the image, where on this map is necessary locate the most populated zones in order to obtain an approximation to the vanishing points.

Regularly there are three orthogonal vanishing points, it depends of the figure position on scene, sometimes a good approximation to these points could be difficult because there are other lines that introduce interferences.

The algorithm was developed and also a graphic interface to visualize the algorithm steps easily, see fig. 1.1.

Algorithm steps:

- Line detection,

- Line intersections,
- Intersections clustering.

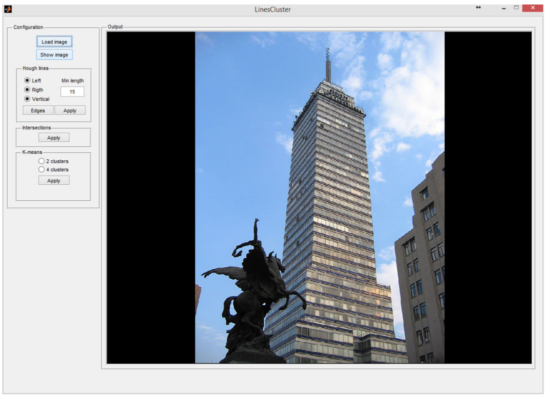


Figure 1.1.- Developed GUI with an image loaded, ready to process it.

Once image was loaded, the edges map is computed. Using the edges map and the Hough transformation, the lines of the image are obtained, Fig. 1.2.

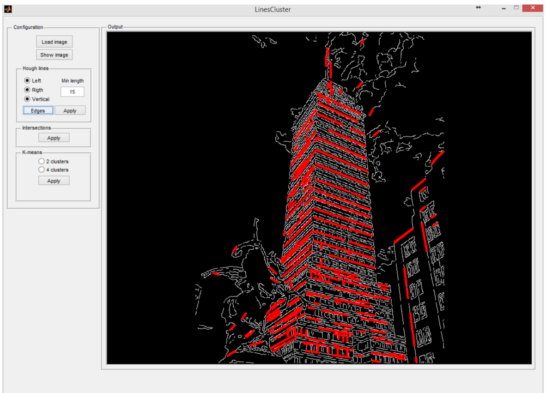


Figure 1.2.- Lines plotted over the edges map.

Note that the vertical lines are minimum, so we need to improve the vertical line detection, it can be achieved through of the study of the behavior of the Hough transform with different parameters.
The next step is the computing of the intersections of the combinations of all the lines found on the image, Fig.1.3.

At this point we have a plane with line intersections, some intersections are very far from the image, and therefore these intersections are discarded.
As we can see on the Figure 1.3 the concentration of points is near of the real vanishing points, because the meeting point of the lines tends to this point.

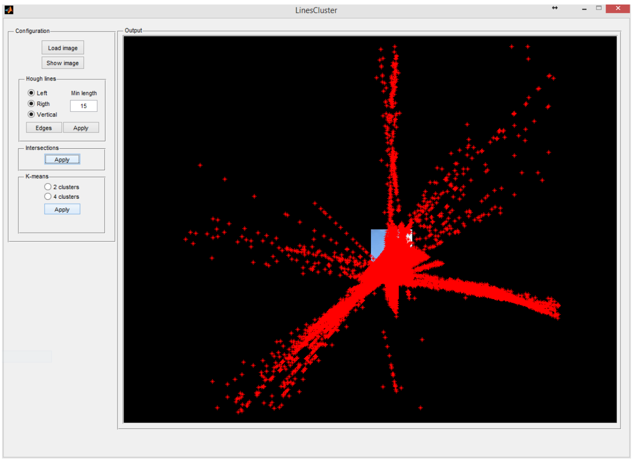


Figure 1.3.- Line intersections plotted over the original image.

The next step is K-Means clustering in order to obtain points to get an approximation to the vanishing points.

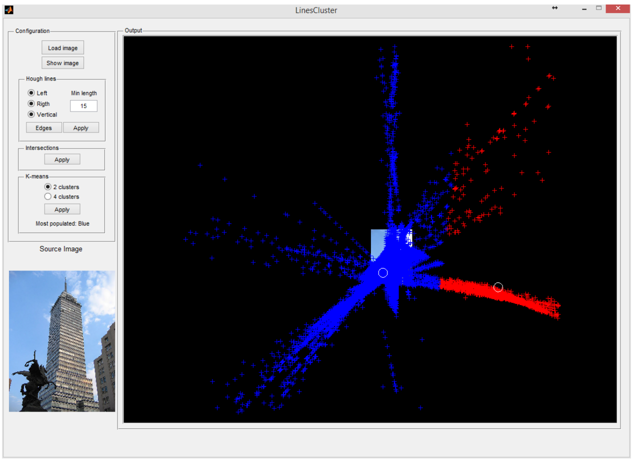


Figure 1.4.- Division in two clusters.

On Figure 1.4 we can see that two clusters are not enough, for this type of structures in some cases three vanishing points could be present.

On Figure 1.5 it can be seen a good approximation to the vanishing points using the centroids of each cluster.

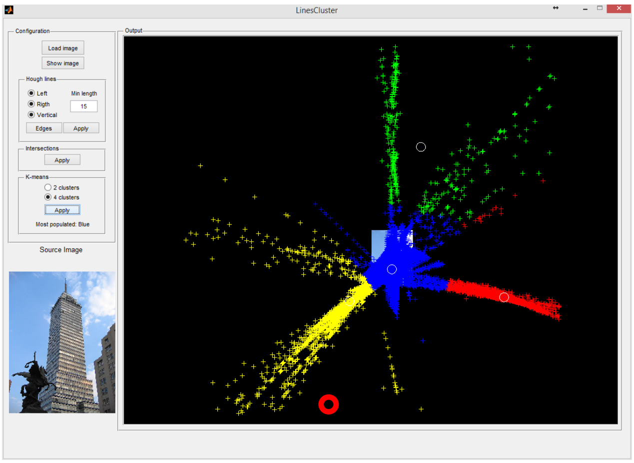


Figure 1.5.- Computed clusters of detected intersections.

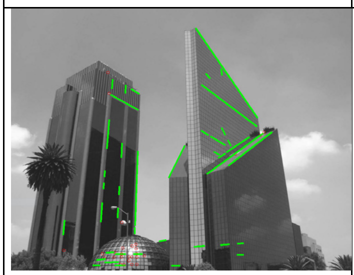
On this step, the relation between lines with the corners was studied, with the aim of detect man-made structures, since on the nature there are not geometric shapes.
The implemented algorithm consists of four steps, on the first, second and third step we only calculate the necessary elements to perform the final processing, on the last step the relations between lines and corners are established.

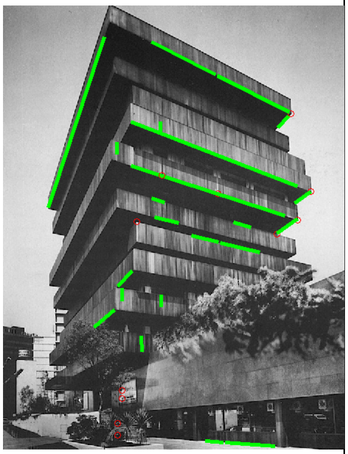
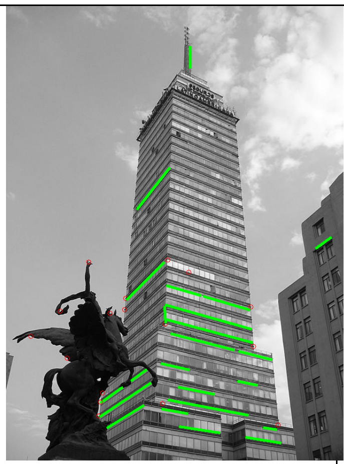
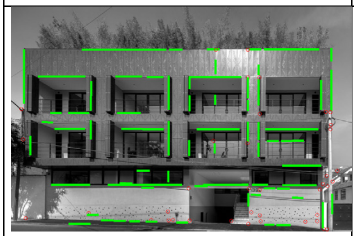
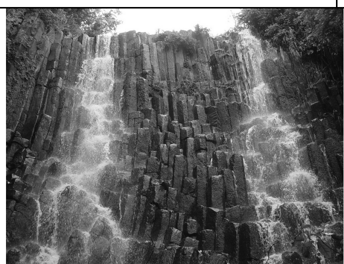
Algorithm steps:

1. Corner detection.
2. Line detection.
3. Lines intersection.
4. Matching between corners and lines.

The following table shows the comparison between images with nature and images with structures made by the man.

Table 1.1

Images	Information	Images	Information
	Corners:74 Intersections:1544 Lines:90 Relations:56		Corners:846 Intersections:2624 Lines:91 Relations:1558

	Corners:132 Intersections:1899 Lines:87 Relations:18		Corners:105 Intersections:343 Lines:36 Relations:26
	Corners:140 Intersections:1856 Lines:90 Relations:506		Corners:68 Intersections: 46 Lines: 26 Relations: 0
	Corners: 59 Intersections: 2 Lines: 10 Relations: 0		Corners: 225 Intersections: 3 Lines: 3 Relations: 0
	Corners: 94 Intersections: 0 Lines: 0 Relations: 0		Corners: 85 Intersections: 0 Lines: 2 Relations: 0

E.2 TEXTURE ANALYSIS

The three principal approaches used in image processing to describe the texture region are statistical, structural [E.1.4] and spectral [E.1.5]. These approaches have been used in many applications where we needed matching or segment region with different texture. The principle in segmentation is to have a texture recognized in an image. This image may contain several different textures. The Local Binary Pattern (LBP) method is a very effective and no parametric texture descriptor. It is invariant to monotonic gray level changes. It was first introduced by Ojala and Pietikäinen [E.1.5,E.1.6].

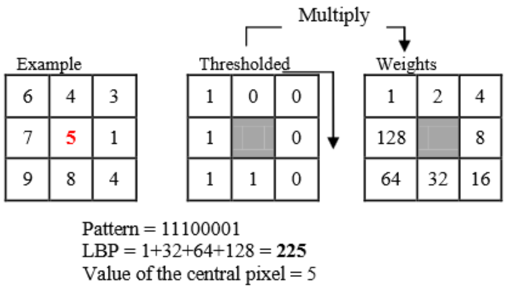
The method is defined as:

$$\text{LBP}_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c)2^p.$$

Where gc is the value of central pixel and gp is the value of pixel adyacent to central. The function s(x) is defined by:

$$s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}.$$

An example of use of LBP is show in the following figure:



A histogram with occurrence of this pattern is known has feature of LBP method. The histogram of the uniform patterns LBP (P, R) on the image or on a region of the image is a powerful feature of texture [E.1.7]. Obtaining the histogram and its statistical occurrences in structural elements are combined considered micro-structures like edges, lines, flat regions and points whose distribution is estimated by the histogram [E.1.8].

Regarding **Mexican architectural structures**, one of the essential issues is the accuracy in detecting edges. The main contribution was to introduce new configuration masks to better approach the partial derivatives and an adaptation stage for the maximum suppression window. This yields better corner candidates. The evaluation database has 26,830 images obtained from 32 segmented videos of the Mex-Culture database. Each image is described by information regarding corners, lines, intersections and corners relationships. The detection rate of architectural structures for the database was 89%.

F Conclusion and Perspectives

Nowadays the Content-based video retrieval techniques are a very important research field to manage and index multimedia databases. Thus, several visual descriptors has been proposed in the last years, which in many cases are dependent on the application as well as on the multimedia database where are applied. In the work carried out, the experimental results showed the accuracy of the proposed fast CBVR techniques applied to a database of Mexican Culture videos which are captured in highly diverse environmental conditions and with different acquisition devices. Furthermore, within Mex-Culture project, for image and video content, specific detectors were devised from differents descriptors mentioned in

ED1.1 report, generating elements of nature, human figures and Mexican architectural structures. The work carried out for video and image processing in the project has been implemented in the architecture of the Mex-Culture platform. This is described in ED3.3, ED4.6 and ED4.8 raports as well as and in the articles written on the subject.

G References

[ED1-1] Mid-term report on scalable visual descriptors tools. Mex-Culture project, 2014. ANR-CONACYT.

[C.1.1] Broomhead, D.S., Lowe, D.: Multivariable Functional Interpolation and Adaptive Network. Complex Systems, 2, 321–355. (1988).

[D.1.1] De la O Torres Saúl, Martínez Nuño J. Alfredo, García Vázquez Mireya S, Hernández García Rosaura. Búsqueda de personas mediante el uso de detección de piel en una secuencia de video. Congreso Internacional en Ingeniería Electrónica. Mem. Electro 2014, Vol.36, pp. 255-260. Chihuahua, Chih. México. October 2014.

[E.1.1] A. Almansa, A. Desolneux, y S. Vamech. Vanishing point detection without any a priori information, IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, núm. 4, pp. 502–507, abr. 2003.

[E.1.2] C. Harris y M. Stephens, A combined corner and edge detector. In Alvey vision conference, 1988, vol. 15, p. 50.

[E.1.3] W. Richards y A. Polit, “Texture matching”, Kybernetik, vol. 16, núm. 3, pp. 155–162, sep. 1974.

[E.1.4] R. M. Haralick, “Statistical and structural approaches to texture”, Proc. IEEE, vol. 67, núm. 5, pp. 786–804, may 1979.

[E.1.5] A. C. Bovik, M. Clark, y W. S. Geisler, “Multichannel texture analysis using localized spatial filters”, IEEE Trans. Pattern Anal. Mach. Intell., vol. 12, núm. 1, pp. 55–73, ene. 1990.

[E.1.6] T. Ojala, M. Pietikainen, y D. Harwood, “Performance evaluation of texture measures with classification based on Kullback discrimination of distributions”, en , Proceedings of the 12th IAPR International Conference on Pattern Recognition, 1994. Vol. 1 - Conference A: Computer Vision amp; Image Processing, 1994, vol. 1, pp. 582–585 vol.1.

[E.1.7] M. Pietikäinen, T. Ojala, y Z. Xu, “Rotation-invariant texture classification using feature distributions”, Pattern Recognit., vol. 33, pp. 43–52, 2000.

[E.1.8] T. Ojala, M. Pietikainen, y T. Maenpaa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns”, IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, núm. 7, pp. 971–987, jul. 2002.

[E.1.9] I. Oulu, The Local Binary Pattern Approach to Texture Analysis – Extensions and Applications. 2003.

[ED3-3] Final report on summarization and scalable search. Mex-Culture project, 2015. ANR-CONACYT.

[ED4-6] Report on multimedia platform. Mex-Culture project, 2015. ANR-CONACYT.

[ED4-8] Final report on multimedia platform. Mex-Culture project, 2016. ANR-CONACYT.