AGENCE NATIONALE DE LA RECHERCHE ANR	Compte-rendu de fin de projet	

Projet ANR-11-IS-001

MEX-CULTURE

Final Report

Programme ANR Blanc International II 2011

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)	Instituto Politécnico Nacional
Project coordinator (if applicable)	Michel Crucianu
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://mexculture.cnam.fr

Author of this rapport	
Title, first name, surname	Prof. Michel CRUCIANU Prof. Jenny BENOIS-PINEAU Prof. Henri NICOLAS Prof. Mireya Saraí GARCIA-VAZQUEZ Dr. Alejandro RAMIREZ-ACOSTA
Telephone	(+33) 1 40 27 24 58
E-mail	Michel.Crucianu@cnam.fr
Date of writing	30/12/2015
Time period covered by this report	01/01/2012 – 31/03/2016

Liste des partenaires présents à la fin du projet (société/organisme et responsable scientifique)	CEDRIC – Cnam (Michel CRUCIANU) LABRI (Jenny BENOIS-PINEAU) IPN (Mireya GARCÍA-VÁZQUEZ) UNAM (Francisco GARCÍA-UGALDE)

B Resume consolide public

B.1 Instructions pour les resumes consolides publics

B.2 Resume consolide public en français

Structuration et exploration interactive de collections culturelles audio-visuelles

Nouvelles méthodes d'indexation, structuration et recherche par le contenu dans de grandes collections multimédia

L'héritage culturel a un rôle majeur dans la promotion de la diversité dans un monde globalisé, il est donc très important de rendre ces contenus facilement accessibles à un large public. De grandes collections de contenus culturels doivent être indexées et les utilisateurs doivent disposer d'outils leur permettant un accès facile et rapide aux données multimédia pour la recherche (suivant des critères multiples) et la visualisation du contenu. Le projet Mex-Culture cherche à mettre au point de nouvelles méthodes automatiques pour le traitement à large échelle de données multimédia. Ces méthodes concernent l'indexation vidéo scalable, l'indexation audio en utilisant des descripteurs issus de la reconnaissance de la parole et de l'identification du locuteur, l'indexation pluri-modale (image, vidéo avec audio et parole), ainsi qu'une recherche par le contenu qui passe à l'échelle. Les techniques développées sont mises en œuvre dans une plate-forme commune de développement et, dans le cadre du projet, appliquées en premier à des contenus culturels mexicains issus des grandes collections de Canal Once, de FONOTECA NACIONAL et de la bibliothèque de programmes vidéo de l'UNAM (TVUNAM), CIESAS, INAH-DL.

Description multiple par le contenu, apprentissage statistique et recherche efficace dans une grande collection

Les travaux ont abordé d'abord de nouveaux descripteurs visuels combinant des informations locales et globales de couleur et de formes. Ensuite, des outils de détection et d'indexation basés sur des descripteurs multiples et des méthodes d'apprentissage statistique ont été mis au point. Ils permettent d'identifier des concepts importants dans les contenus culturels mexicains, comme des éléments de la nature, des figures humaines et des structures architecturales. Ils permettent également de détecter cinq classes de contenu audio, quatre langues indigènes et de séparer musique et parole. Des descriptions vidéo et audio ont été employées conjointement, avec des méthodes d'apprentissage statistique, pour obtenir des résumés vidéo. Des détecteurs d'actions ont aussi été mis au point, basés sur des descripteurs compacts de mouvement et combinant plusieurs techniques d'apprentissage statistique. Nous nous sommes également intéressés au passage à l'échelle des résumés vidéo et de la détection d'actions. Une plate-forme commune a été développée, avec une architecture ouverte basée sur les services web, afin de faciliter le développement indépendant de composantes complémentaire pour la description et la recherche de données multimédia.

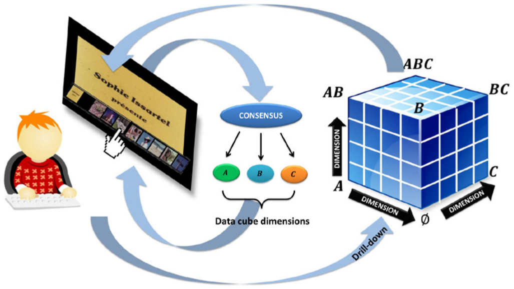
Résultats majeurs du projet

Les efforts des participants au projet ont produit (i) une grande base de contenus culturels mexicains, incluant des corpus annotés pour les différentes tâches du projet ; (ii) de nouvelles méthodes qui passent à l'échelle pour indexer, résumer et rechercher de grandes collections de contenus culturels multimédia ; (iii) de nouvelles méthodes trans-modales pour la structuration et l'exploration de contenus audio-visuels ; (iv) une plate-forme commune ouverte pour l'indexation et la recherche de contenus culturels multimédia. Le projet a renforcé la collaboration entre partenaires français et mexicains, ainsi qu'avec des institutions culturelles mexicaines.

Production scientifique

Les résultats du projet ont déjà fait l'objet de 16 communications dans des conférences ou workshops internationaux avec actes, ainsi que de 2 articles dans les journaux internationaux *Multimedia Tools and Applications* (Springer) et respectivement *IEEE Transactions on Circuits and Systems for Video Technology*. Sur ces publications, 16 sont multi-partenaires dont 13 qui regroupent des partenaires français et mexicains.

Illustration



Résumé vidéo sous la forme d'un cube de données

Informations factuelles

Mex-Culture est un projet de recherche fondamentale franco-mexicain coordonné par le Cnam (Paris). Il associe aussi le LaBRI (Bordeaux), l'INA (Bry-surMarne) et deux partenaires mexicains, l'IPN (México) et l'UNAM (México). Le projet a commencé en février 2012 et a duré 50 mois. Les partenaires français ont bénéficié d'une aide ANR de 276 702 € pour un coût global de l'ordre de 991 500 €.

B.3 Resume consolide public en anglais

Structuring and interactive exploration of audio-visual cultural collections

Novel methods for indexing, structuring and content-based retrieval from large multimedia collections

Given the importance of cultural heritage content in promoting diversity in a globalized world, making this content readily available to a broad audience is a critical issue. Large volumes of such content must be indexed and users must be provided with means for a fast and easy access to the multimedia information, making them able to browse (according to multiple criteria) and visualize desirable content stored in the archives. The Mex-Culture project aims to devise novel automated methods for large-scale processing and indexing of multimedia content. These methods concern scalable video indexing, audio indexing using descriptors issued from speech recognition and speaker identification, cross-media indexing (image, video plus audio and speech), as well as scalable search and retrieval. The resulting techniques are implemented in a common development platform and, within the project, primarily applied to Mexican cultural content from the large databases of Canal Once, of FONOTECA NACIONAL and of the Video library of the UNAM (TVUNAM), CIESAS, INAH-DL.

Multiple content description, machine learning and efficient retrieval from large multimedia collections

The efforts first focused on new visual descriptors combining local and global color and shape information. Then, specific detectors and indexing tools, relying on multiple descriptions and machine learning, were devised. They allow to identify important concepts in audio-visual cultural Mexican content, including elements of nature, human figures and architectural structures. They also allow to detect five classes of audio content, four indigenous languages and separate music and speech. Video and audio descriptors were jointly employed, together with machine learning, for video segmentation and summarization. Action detectors were also devised, relying on compact movement descriptions and combining several machine learning techniques. Specific work addressed the scalability issue for both video summarization and action detection in videos, i.e. making the proposed methods efficiently applicable to large video databases. The common platform has an open architecture based on web services in order to facilitate the independent development of complementary content description and retrieval components.

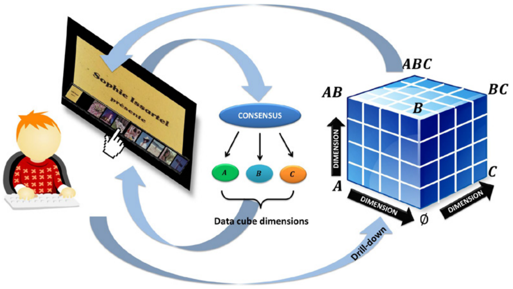
Major results of the project

The work within the project resulted in (i) the creation of a large database of Mexican cultural content, including specific annotated corpora for different tasks of the project; (ii) novel, scalable methods for indexing, summarization and retrieval in large cultural archives; (iii) novel cross-modal methods for the structuring and exploration of audio-visual content; (iv) a common open platform for indexing and retrieval of multimedia cultural content. The project reinforced the collaboration between French and Mexican partners, as well as with cultural institutions in México.

Scientific results

The results obtained during the project were published in 16 papers at international conferences or workshops and in 2 papers in the international journals *Multimedia Tools and Applications* (Springer) and *IEEE Transactions on Circuits and Systems for Video Technology*. Among these publications, 16 are multi-partner and 13 of these are co-authored by both Mexican and French partners.

Illustration



Video summary as a data cube

Information

Mex-Culture is a French-Mexican fundamental research project coordinated by the Cnam (Paris). The other participants are the LaBRI (Bordeaux), INA (Bry-surMarne) and two Mexican partners, IPN (México) and UNAM (México). The project started in February 2011 and went on for 46 months. The French partners received 276,702 € of financial support from the ANR for an overall cost of about 991,500 €.

C Scientific report

The following presentation is a brief account of the research performed in the project. Further details regarding visual and audio content description can be found in ID1.2, ED1.1 and ID2.2, ED2.1, while details regarding content summarization and scalable search are given in ED3.3.

Mémoire scientifique confidentiel : non

C.1 CHALLENGES, STATE OF THE ART

(numerical references correspond to publications of the Mex-Culture project, while alphanumerical references concern external literature)

Given the importance of cultural heritage content in promoting diversity in a globalized world, making this content quickly available to a broad audience is a critical issue. Large volumes of such content must be indexed and users should be provided with means for a fast and easy access to the multimedia information, making them able to browse (according to multiple criteria) and visualize desired content stored in the archives. The research activities in this project aimed to bring new contributions in three areas: content description, content summarization and scalable content-based search.

Unlike in other European projects on large scale multimedia content search engines where the reference corpora already existed, in Mex-Culture it was necessary to create the whole infrastructure (Mexican cultural material in appropriate digital format, information quality, information copyright, database software design, DB hardware design, DB management, DB site, DB connection, DB characteristics) for the needs of the project. While being a big challenge for the project's members, this was also an important and interesting innovation. Indeed, in México it was impossible to find a corpus of 100,000 hours of digitized video, with acceptable quality, in accordance with the international standards of preservation and diffusion, documented, cataloged and covering Mexican architecture, traditional dances of México, nature of México, traditional events such as bullfighting, etc. It is important to mention that the Mex-Culture corpus construction was not considered as a task in the project, but for the project's Mexican participants it represented an important activity with high resources consumption. Also, the overall corpus construction process involved technical setup in terms of devices and software used for database management and annotation. Specific concepts design, file identifications design and annotation scheme were other important activities. Personnel for making annotations was needed and this was also a significant time investment. Cultural aspects were analyzed to devise the specific concepts for the Mex-Culture database [14,15].

C**ontent description**. In the framework of the project, new descriptors for digital audio-visual content were proposed. Audio-visual documents and digital photos are dominant in the heterogeneous pool of cultural multimedia content. Consequently, content-based descriptors of **images** and/or **key-frames** still represent an active research field. Unlike the local descriptors (the most popular being SIFT [Low04] and SURF [WTG08]), the global descriptors from MPEG 7 standard such as DCD and CLD [SM02] allow for color characterization in the whole image or video frame. The goal of visual content description in the project was to combine in an efficient way locality, color and shape for CBIR or CVIR by key-frame.

C**ontent summarization**. Video documentaries are one way to capture the cultural heritage of a country and they can be used for the preservation and dissemination of the culture. Large volumes of such content as well as their large duration enhance the necessity of developing a fast, easy and multidimensional access to them. The task of video summarization is a way of providing compact representations of video contents by extracting their "most relevant" information. The development of the information retrieval and browsing fields in large amounts of multimedia data encouraged the design of video summarization approaches [BBL06, JHC10, ALT13] and a specific task was run in the context of the TRECVID evaluation campaigns. Video summary generation still remains an important open problem, as testified by recent works in multimedia research [GCG14]. Generally speaking, video summarization approaches mainly consist in grouping similar video segments on the basis of continuous audio channel [KML12]. They often rely on data analysis techniques such as clustering, supervised learning, etc. The strong requirements of those applications in terms of scale, time response and high-dimensional information make the **scalability** a very challenging problem. Scalability can be seen as the ability to deal with large amounts of data efficiently. Another interpretation comes from multi-view data representation and means that data can be described in a coarse-to-fine manner: this is how we understand the scalable video summarization here. A scalable video summary allows a user to browse abstracted video contents in a progressive manner.

Scalable content-based search. Content-based search is another key functionality in providing access to multimedia content. Within the project, we focused on content-based search for images, audio and actions in video. For image retrieval, global features especially concerning color are widely used [Liu96, How00, Ste02] because color is visually very relevant and is insensitive to image translation, scaling and rotation.

Detection and localization of human actions in videos is challenging because of the complexity and variability of human motions, but also because of the large amount of video data to be searched. It remains an open problem, despite intensive research during the past decade. Three semantic levels are typically considered. An *atomic* action is simply a short coherent elementary movement such as "raise hands" or "move leg". At this level, research mainly focuses on modeling such actions as statistical processes [HS12] or as time series [ZTH08]. At the *intermediate* level, an action is composed of a series of atomic parts and can vary in complexity, e.g. from "smoking" to "pole vaulting". Most of the recent research considers this level and the focus is on finding ways to aggregate atomic descriptions. Finally, at a *higher* semantic level, interest is in "events" that group actions into classes having high variability in terms of both atomic components and temporal organization (e.g. "making a sandwich" in the TRECVID MED challenge [OAM13]). Complex background, variability in point of view, occlusions and low video quality are challenges for action detection in video. Actions in video are modeled using either global descriptions of spatiotemporal volumes of the video [BGS05, LP07, CAR08, TCL12] or sets of local features describing spatiotemporal patches [Lap05, DRG05, DLS09, GHS11, OVS13, OVS14, SJX14]. With local features, modeling relies on their statistical distribution over a volume of the video. Volumetric methods are particularly useful for precise spatiotemporal localization but expensive when used at a large scale. Methods based on local features recognize actions through the statistics of sets of descriptors of small video regions, not necessarily linked to body parts or image coordinates. The advantage is in avoiding the (error-prone) segmentation of the human from the background, as well as the computation of a costly description of a full video volume.

A user may wish to access previously found classes but also be able to define novel action classes by providing examples. With a large database, it can be prohibitively time consuming to perform a new exhaustive scan of the entire database every time a new class detector is built. It is then necessary to devise methods supporting the *scalable* application of a detector to the data, i.e. methods that are *sublinear* in the size of the database. Action detection on large scale datasets was not specifically addressed in the literature. Furthermore, while the scalability of query-by-example was thoroughly considered in the multimedia retrieval literature, there is comparatively little work on the scalability of what we call *query-by-detector*.

Several proposals focus on sublinear methods that aim to find the data points whose image in feature space is close to the normal vector to the SVM hyperplane [PC06, KKH14, KYD14].

C.2 SCIENTIFIC AND TECHNICAL APPROACH

Content description. 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. 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, then feeding these descriptors into neural networks [BL88, Sal15]. Fig. 1 shows the process of detection of elements of nature.

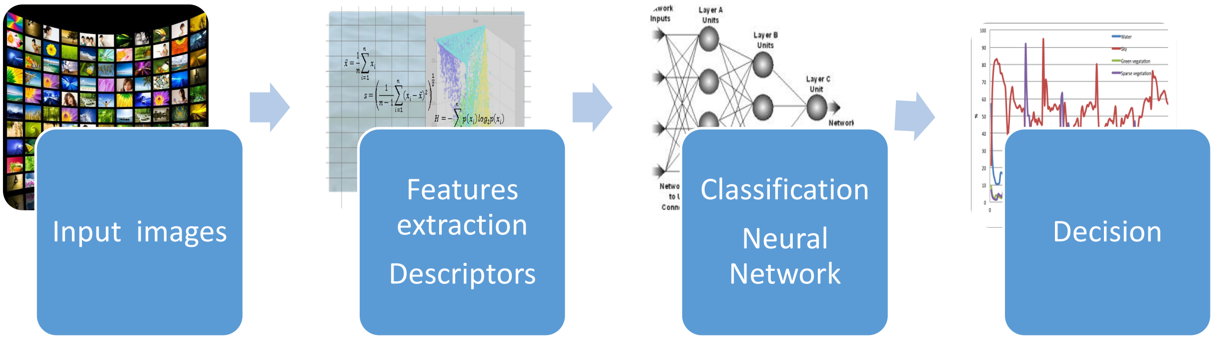


Figure 1: Detection elements of the nature.

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 [DTM14]. Regarding detection, our main 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) [ADV03] 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 [ADV03]. With this method, we are looking for the relations between lines and corners [HS88], using the Canny edge detector and the Hough transform.

The **audio descriptors** for speech / music classification in cultural content are Perceptual Linear Prediction (PLP) coefficients that are widely employed in speech processing. Audio/visual approaches are also used by combining audio (MFCC descriptors) and visual parameters (chroma vectors, dominant hue and lightness descriptors) to enrich the audio-visual content description. The proposed descriptors are presented in the deliverables ID1.2, ED1.1 and ID2.2, ED2.1. Within Mex-Culture, **specific detectors** were also devised for audio content classes and indigenous languages. Speech recognition was also employed. Fonoteca Nacional México [FNM15] classifies its audio content in five classes: sound Art, Music, sound Landscape, Radio, and Voice. In general, each class is composed of a mixture of audio signals: music with different types of genres, voice only, voice with music, voice with different sounds in the backgrounds, nature sounds, animals sounds. We obtain a **classification of audio files with regard to these five classes**. We then do a deeper analysis of each class in order to generate new descriptors that will provide more detailed information on the content. For this identification of five classes we are inspired by work on acoustic landmarks [Ste02, Liu96]. We implemented descriptors that correspond to concentrated energy localized in time and frequency (ED2.1) and use them for landmark-based sound recognition. Another contribution is to **identify indigenous languages** (Maya – Yucatec, Nahuatl – Central, Otomi variant Hñahñu, PaiPai) based on speech fingerprinting descriptors that are the onsets formed into pairs. They are parameterized by the frequencies of the peaks and the time in between them. These descriptors are quantized to give a relatively large number of distinct landmark-based speech fingerprinting hashes (ED2.1). For **speech recognition**, we employed the hidden Markov model (HMM) [Rab89] to implement the isolated word speech recognition system. The information extracted from the audio signal concerns MFCC and, to increase the information of the human perception, the first and second time derivatives are calculated.

Content summarization. Here, our main contribution is to provide a scalable video summarization approach inspired by On-Line Analytical Processing (OLAP) cube operations [GCB97]. The idea is borrowed from hierarchical information retrieval frameworks, which have become particularly popular in Multimedia archives [BPS11]. The data cube (or hypercube) concept has been proposed to facilitate the navigation of the user through multidimensional spaces, where each move corresponds to a query using some combination of the dimensions. In this work, we consider different descriptors and embed them into a consensus clustering framework allowing to partition the data cube according to a cross-modal feature space. This approach is evaluated on a sample of the INA cultural video corpus. To achieve the scalable video summarization, we follow the methodology illustrated in Fig. 2 below, which mainly consists of two stages, called video summary and scalability.

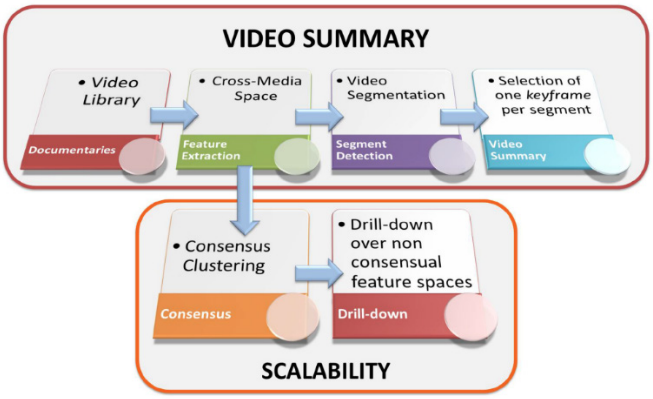


Figure 2: Scalable video summary construction

The approach we consider for video segmentation relies on the clustering of early-fused features for high dimensional feature spaces, and on the consensus clustering paradigm for low-dimensional feature spaces (see the Consensus block in the figure). Consensus clustering consists in merging different clusterings performed over different dimensions of the description space. Our goal is to ensure a scalable navigation in a video summary. We use the OLAP data cube to model a video document in a cross-media description space, composed of audio and video feature dimensions [GCB97]. Such a model allows the user to navigate into the clusters obtained according to his preferences. We materialize this by the "drill down" block in the figure above.

Scalable content-based search. Content-based Image Retrieval has been a very intensively researched are since the last two decade. While the general trend in CBIR has moved toward supervised classification schemes instead of metric-based comparison (1-NN) search, there still remains the place for the descriptor development, specifically by fusion of local and global features. The comparison of **images** in the composed description space is then performed by combination of metrics computed in each subspace. This was the approach proposed in the project for key-frame-based video or image search [4].

For **audio** search, the project focused on detection of speech/music. The LaBRI Speech/Music detection system is built using data from the ESTER evaluation campaigns [GGG06]. The features employed are PLP coefficients which are widely used in speech processing. We train GMM models for each class. Viterbi decoding is used during the test phase. This method has proved to be efficient on the data used in the ESTER evaluation campaign, especially for speech detection (F-measure of 0.93). The same method has been applied to the data of the project. The audio files have been annotated manually for evaluation purposes by several annotators. These annotations were provided by the Mexican partners. There were 5 annotators who annotated a total of 40 hours of audio data (64 files).

We made four contributions regarding **scalable retrieval and localization of *intermediate*-level actions** that distinguish our system from previous ones: (i) to take advantage of the temporal information, we represent actions as time series and compare them using the Global Alignment (GA) kernel; (2) to find a better balance between efficiency and effectiveness, we propose a cascaded approach that employs aggregated data in the first level and frame sequence comparison (with the GA kernel) at the second stage; (3) to improve time series comparisons with the GA kernel, we introduce a novel feature selection method for sparse multivariate time series; (4) we introduce two novel methods for scalable retrieval (both based on LSH), one of which is sublinear in the size of the database. The block diagram of the proposed system is shown in Fig. 3 below.

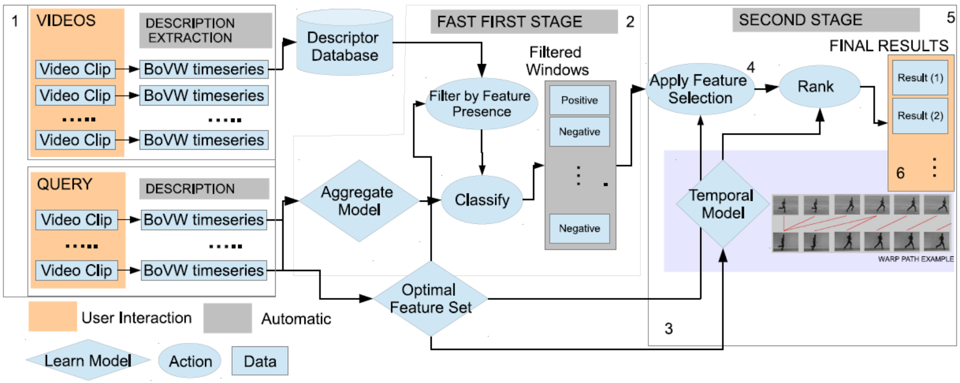


Figure 3: Block-diagram of scalable retrieval and localization of actions in video documents

C.3 RESULTS

Content description. Image content descriptors were developed for content-based video retrieval on the basis of key-frames. Each key-frame is considered as an image in the overall database. A dominant Color Correllogram Descriptor (DCCD), combining local and global information was proposed first [4]. Then a new visual descriptor, which is a linear combination of DCCD and the shape descriptor Pyramidal Histogram of Oriented Gradients (PHOG). The proposed descriptors have shown a 9% improvement (on average) of Average Recall Rate (ARR) and Average Precision Rate (APR) over the SOA in a CBIR task. 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. The evaluation of the **detection of human figures** based on skin color information was made with images extracted from Mex-Culture videos databases. The results show a detection rate of 81%. To increase this accuracy we proposed some improvements in the algorithms. 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%.

In **speech /music** detection the performances obtained are a little bit lower than what we had on the ESTER database, with F-measure for speech detection varying from 0.82 to 0.88 according to the annotator (between 11% and 13% of error). Unfortunately, the results for music detection are not very satisfactory, with F-measures ranging from 0.56 to 0.67 according to the annotator, corresponding to error rates between 34% and 40%. We aim at estimating the temporal boundaries of music pieces relying on the assumed homogeneity of their musical and visual properties. We consider an unsupervised approach based on the generalized likelihood ratio to evaluate the presence of statistical breakdowns of MFCCs, Chroma vectors, dominant Hue and Lightness over time. An evaluation of this approach on 15 manually annotated concert streams shows the advantage of combining tonal content features to timbral ones, and a modest impact from the joint use of visual features in boundary estimation [2]. Furthermore, we proposed to analyze the structural regularities from the **audio and video** streams of TV programs and explore their potential for the classification of videos into program collections. Our approach is based on the spectral analysis of distance matrices representing the short and long-term dependencies within the audio and visual modalities of a video. We propose to compare two videos by their respective spectral features. We appreciate the benefits brought by the two modalities on the performances in the context of a K-nearest neighbor classification, and we tested our approach in the context of an unsupervised clustering algorithm. These evaluations are performed on two datasets of French and Italian TV programs [16]. It should also be noted that the development of audio-visual indexing tools required a strong annotation effort for completely new collections of cultural audio-visual content. An annotation methodology was proposed and OpenSource annotation tools were used together with ad-hoc development of ergonomical Matlab-based

SW tools [14], [15]. This allowed for reduction of errors of individual human annotators and robust ground-truthing of the content.

For Mexican **audio content identification**, the evaluation database comprises 3.4 hours of audio. The integration of the segmentation into voice / music / silence provides useful information for processing the audio content. An identification accuracy of 91% was obtained. To improve effectiveness, a more robust segmentation is required. For **Mexican indigenous languages identification**, the evaluation database comprises 1:25:41 hours of speech files. We obtained an identification accuracy of 93%.

Content summarization. The summarization approach we proposed provides a customized access to several versions of different levels of detail of a video summary in cross-media space. The proposed video summary relies on nonconsensual feature spaces to achieve scalability. We have performed an evaluation of the proposed method with regard to video summaries obtained by a random selection of clusters with arbitrary abstraction with a constant time step and summaries obtained from humans. The method was successfully applied to generic video content without a clearly defined structure, such as cultural documentaries. These results were published in [11]. See also ED3.3.

Scalable content-based search. One contribution of the project was the **creation of MEXaction**, **a new challenging dataset** for action detection and localization. Unlike publicly available datasets like KTH or MSR2, containing rather simple actions such as walking or hand-waving in a reasonably cluttered environment, MEXaction contains excerpts from real-life cultural Mexican content (Corrida). The dataset containing 117 videos (77 hours) was mastered from videos from INA (France) and Canal Once (Mexico) archives, annotated for the ground truth and made publicly available on http://mexculture.cnam.fr/xwiki/bin/view/Datasets/Mex+action+dataset.

Our proposals for **scalable action detection and localization** was evaluated on *MEXaction* as well as on a few existing (smaller) datasets like KTH, MSR2. A substantial gain was obtained by using the novel feature selection method for the time alignment of sequences with the GA kernel, together with a large reduction in the amount of data that has to be loaded from the disk, contributing to the scalability of the solution. The proposed method achieves better performance than the state of the art while having lower memory requirements. These results were published in [12]. Also, our novel LSH-based method can approach the effectiveness of exact exhaustive search while being much more efficient since it only examines a fraction of the data. The method is not dependent on kernel type and parameters, nor on database size. These results were published in [9]. See also ED3.3.

During the project time frame, a new content representation approach based on deep learning with CNN has emerged both in computer vision and multimedia communities. At Cnam and LABRI the study of these tools is conducted in other research projects.

C.4 Application of results

The methods for scalable visual summary construction were developed as a well-structured SW package that has been integrated into the Mex-Culture Multimedia Platform. Furthermore, the “scalable summary service” turned to be attractive for the big regional project Gaiard aiming at the creation of digital content platforms and services; the negotiations are ongoing.

The Mex-Culture platform will be tested by the National Sound Archive of Mexico, for identifying and finding audio content. The goal is to have an automatic classification of audio files with regard to the five classes of the cataloging of the Fonoteca Nacional México. The Mex-Culture platform will also be tested by the Department of Linguistics of the INAH (México) for the identification and searching of indigenous languages.

C.5 Discussion

An important difficulty for the project was the 11 months delay in the financing of the Mexican partners. The French partners had to begin their work with content descriptors from the literature instead of descriptors developed in the project. A second important difficulty was the limited participation of UNAM to the initial work programme. A major part of their activities within the project had to be transferred to IPN, resulting in additional delays. For this reason, the CONACYT decided to give a 6 months extension to IPN to finalize the current activities and associated deliverables (ED1.3, 1.4, 2.3, 2.4, 4.8, 4.9). The fact that the multimedia platform was not completed before the project ended on the French side prevented INA from performing the user evaluation work (deliverable ED4.10 was canceled). Nevertheless, INA provided instructions regarding user evaluation to IPN and part of this evaluation work will be performed outside the project by IPN, Fonoteca Nacional México and the Department of Linguistics of the INAH.

In spite of these difficulties, we believe that the main goals of the project were attained: (i) creation of a large database of Mexican cultural content, including specific corpora for different tasks of the project; (ii) methods for indexing and retrieval in large cultural archives were developed; (iii) scalability issues in both summarization and retrieval were addressed; (iv) cross-modal methods for the structuring and exploration of audio-visual content were put forward; (v) a common platform for indexing and retrieval of multimedia content.

C.6 Conclusions

In the framework of the Mex-Culture project, the following advances in the SOA have been proposed:

- four new image descriptors for CBIR/CBVIR and two new audio descriptors for audio classification;
- new tools for the segmentation into speech / music;
- new tools for the detection of specific components in images (or key-frames): elements of nature, human figures, architectural structures;
- new tools for the detection in audio of five classes of Mexican audio content and four indigenous languages;
- an original approach for scalable audio-visual content summarization;
- a scalable approach for action detection and localization in unstructured cultural video content.

To integrate these components, an open architecture using web services was proposed and a demonstrator was developed. Additional tools were also developed or adapted, including video segmentation in shots or audio-visual annotation.

From a research perspective, this project has opened a new research axis in México, regarding cross-modal and scalable indexing of large digital audio-visual collections. French partners have taken advantage of this collaboration via joint work with Mexican researchers. Components developed in this project are of interest for industrial transfer (Gaiard project but not only). New collaborations between French and Mexican partners are started. A visit took place on February 2016 of Pr. Mireya García-Vázquez with two Master students from IPN-

CITEDI to LABRI, supported by CONACYT. Also, Pr. Jenny Benoit will visit IPN-CITEDI in August 2016. New international agreement between IPN-CITEDI and Bordeaux University are underway to participate in dual academic formation at master and doctoral degree with support of CONACYT.

C.7 References

[ADV03] 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.

[ALT13] Almeida, J., Leite, N.J., da S. Torres, R.: Online video summarization in compressed domain. Journal of Visual Communication and Image Representation 24, 729-738 (2013)

[BBL06] Benini, S., Bianchetti, A., Leonardi, R., Migliorati, P.: Extraction of Significant Video Summaries by Dendrogram Analysis. In: Proceedings of the International Conference on Image processing (ICIP), pp. 133-136 (2006).

[BGS05] Moshe Blank, Lena Gorelick, Eli Shechtman, Michal Irani, and Ronen Basri. Actions as space-time shapes. In The Tenth IEEE International Conference on Computer Vision (ICCV'05), pages 1395-1402, 2005.

[BPS11] Bartolini, I., Patella, M., Stromei, G.: The Windsurf Library for the Efficient Retrieval of Multimedia Hierarchical Data. In: Proceedings of ACM Special Interest Group on Multimedia (SIGMM), pp. 139-148 (2011).

[BL88] Broomhead, D.S., Lowe, D.: Multivariable Functional Interpolation and Adaptive Network. Complex Systems, 2, 321–355. (1988).

[CAR08] Chuohao Yeo, P. Ahammad, K. Ramchandran, and S.S. Sastry. High-speed action recognition and localization in compressed domain videos. Circuits and Systems for Video Technology, IEEE Transactions on, 18 (8): 1006-1015, Aug 2008.

[DTM14] 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.

[DLS09] Olivier Duchenne, Ivan Laptev, and Josef Sivic. Automatic annotation of human actions in video. In Proc. of the Intl. Conf. on Computer Vision (2009), pages 1491-1498, 2009.

[DRG05] P. Dollar, V. Rabaud, G. Cottrell, and S. Belongie. Behavior recognition via sparse spatio-temporal features. In Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005. 2nd Joint IEEE International Workshop on, pages 65-72, Oct 2005.

[FNM15] Fonocata Nacional de México, Conaculta: http://www.fonotecanacional.gob.mx/.

[GGG06] Galliano, S., E. Geoffrois, G. Gravier, J.-F. Bonastre, D. Mostefa and K. Choukri. Corpus description of the ESTER evaluation campaign for the rich transcription of French broadcast news. Proc. Language Evaluation and Resources Conference, 2006.

[GCB97] Gray, J., Chaudhuri, S., Bosworth, A., Layman, A., Reichart, D., Venkatrao, M., Pellow, F., Pirahesh, H.: Data cube: A relational aggregation operator generalizing group-by, cross-tab, and sub totals. Journal of Data Mining and Knowledge Discovery 1(1), 29-53 (1997).

[GCG14] Gong, B., Chao, W.L., Grauman, K., Sha, F.: Diverse Sequential Subset Selection for Supervised Video Summarization. In: Proceedings of the Neural Information Processing Systems Conference (NIPS), pp. 1-9 (2014).

[GHS11] A. Gaidon, Z. Harchaoui, and C. Schmid. Actom Sequence Models for Efficient Action Detection. CVPR 2011 - IEEE Conference on Computer Vision & Pattern Recognition, pages 3201-3208, June 2011.

[How00] Andrew Wilson Howitt. Vowel landmark detection. In Proc. ICSLP, 2000.

[HS88] C. Harris y M. Stephens, A combined corner and edge detector. In Alvey vision conference, 1988, vol. 15, p. 50.

[HS12] M. Hughes and E. B. Sudderth. Nonparametric discovery of activity patterns from video collections. IEEE Computer Vision & Pattern Recognition Workshops}, pages 25-32, June 2012.

[JHC10] Jin, X., Han, J., Cao, L., Luo, J., Ding, B., Lin, C.K.: Visual Cube and n-line analytical processing of images. In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM), pp. 849-858 (2010).

[KKH14] Youngdae Kim, Ilhwan Ko, Wook-Shin Han, and Hwanjo Yu. iKernel: Exact indexing for support vector machines. Information Sciences, 257 (0): 32-53, 2014.

[KML12] Kompatsiaris, Y., Merialdo, B., Lian, S. (eds.): TV Content Analysis: Techniques and Applications. CRC Press (2012).

[KYD14] Arijit Khan, Pouya Yanki, Bojana Dimcheva, and Donald Kossmann. Towards indexing functions: Answering scalar product queries. In Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, SIGMOD'14, pages 241-252, New York, NY, USA, 2014.

[Lap05] I. Laptev. On space-time interest points. Intl. Journal of Computer Vision, 64 (2-3): 107-123, September 2005. ISSN 0920-5691.

[Liu96] Sharlene A. Liu. Landmark detection for distinctive feature-based speech recognition. Journal of the Acoustical Society of America, 100(5):3417-3430, Nov. 1996.

[Low04] Lowe, D. Distinctive image features from scale-invariant keypoints. International Journal on Computer Vision (IJCV), 2(60):91-110, 2004.

[LP07] Ivan Laptev and Patrick Pérez. Retrieving actions in movies. Proc. Int. Conf. on Computer Vision (ICCV'07), pages 1-8, Oct 2007.

[OAM13] P. Over, G. Awad, M. Michel, J. Fiscus, G. Sanders, W. Kraaij, A. F. Smeaton, and G. Quénot. Trecvid 2013 -- an overview of the goals, tasks, data, evaluation mechanisms and metrics. In Proceedings of TRECVID 2013. NIST, USA, 2013.

[OVS13] D. Oneata, J. Verbeek, and C. Schmid. Action and Event Recognition with Fisher Vectors on a Compact Feature Set. In ICCV 2013 - IEEE International Conference on Computer Vision, pages 1817-1824, Sydney, Australia, December 2013. IEEE.

[OVS14] Dan Oneata, Jakob Verbeek, and Cordelia Schmid. Efficient Action Localization with Approximately Normalized Fisher Vectors. In CVPR 2014 - IEEE Conference on Computer Vision & Pattern Recognition, Columbus, OH, United States, June 2014. IEEE.

[PC06] Navneet Panda and E.Y. Chang. KDX: an indexer for support vector machines. Knowledge and Data Engineering, IEEE Transactions on, 18 (6): 748-763, June 2006. ISSN 1041-4347.

[Rab89] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, vol. 77, pp. 257-286, Feb. 1989.

[Sal15] R. Salas. Redes Neuronales Artificiales . http://www.inf.utfsm.cl/~rsalas/Pagina_Investigacion/docs/Apuntes/Redes%20Neuronales%20Artificiales.pdf [May. 20, 2015].

[SJX14] Ling Shao, S. Jones, and Xuelong Li. Efficient search and localization of human actions in video databases. IEEE Trans. on Circuits and Systems for Video Technology, 24 (3): 504-512, March 2014.

[SM02] Salembier, Ph. and B. S. Manjunath, Introduction to MPEG 7: Multimedia Content Description Language, Willey, 2002, 352p.

[Ste02] K. N. Stevens. Toward a model for lexical access based on acoustic landmarks and distinctive features, Acoustical Society of America, vol. 111, no. 4, pp. 1872-1891, 2002.

[TCL12] Y.L. Tian, L. Cao, Z. Liu, and Z. Zhang. Hierarchical filtered motion for action recognition in crowded videos. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 42 (3): 313-323, May 2012.

[WTG08] Willems, G., T. Tuytelaars, and Van Gool L. An efficient dense and scale-invariant spatiotemporal interest point detector. In ECCV, 2008.

[ZTH08] F. Zhou, F. De la Torre Frade, and J. K. Hodgins. Aligned cluster analysis for temporal segmentation of human motion. IEEE Conference on Automatic Face and Gestures Recognition, September 2008.

D DELIVERABLES AND MILESTONES

Date	N°	Nature	Title	Partners (resp.)
T0+11	ED4-1	Web site	Web site of the project	All (Cnam)
T0+16+i , i=1à10	ID4-2	Data	Database available 10% × i, i=1…10 (i is the Month)	IPN, , Cnam (INA)
T0+24	ID4-3	Report	Guidelines of audiovisual documentation practices	Cnam (INA)
T0+28	ED1-1	Report	Mid-term report on scalable Visual descriptors	IPN, UNAM, LABRI
T0+29	ID1-2	Software	Intermediate version of scalable Visual descriptors tools	IPN, UNAM, LABRI
T0+50	ED1-3	Report	Final report on scalable Visual descriptors	IPN, LABRI
T0+50	ED1-4	Software	Final version of scalable Visual descriptors tools	IPN, LABRI
T0+32	ED2-1	Report	Mid-term report on Speech/Audio descriptors	IPN, LABRI
T0+32	ID2-2	Software	Intermediate version of Speech/Audio descriptors tools	IPN, LABRI
T0+50	ED2-3	Report	Final report on Speech/Audio descriptors	IPN, LABRI
T0+50	ED2-4	Software	Final version of Speech/Audio descriptors tools	IPN, LABRI
T0+20	ED3-1	Report	Mid-term report on summarization and scalable search	LABRI, Cnam, IPN
T0+32	ID3-2	Software	Intermediate version of content summarization and search tools	LABRI, Cnam, IPN
T0+50	ED3-3	Report	Final report on on summarization and scalable search	Cnam, LABRI, IPN
T0+50	ED3-4	Software	Final version of content summarization and scalable search tools	Cnam, LABRI, IPN
T0+44	ID4-4	Report	Technical and functional specs multimedia platform	IPN, Cnam, LABRI
Fused with ED4-9	ID4-5	Software	Intermediate version of integration ID1-2, ID2-2 and ID3-2 in multimedia platform	IPN, Cnam, LABRI
T0+46	ID4-6	Report	Report on multimedia platform	IPN, Cnam, LABRI
Fused with ED4-9	ID4-7	Software	Intermediate version of multimedia platform	IPN, Cnam, LABRI
T0+50	ED4-8	Report	Final report on multimedia platform	IPN, Cnam, LABRI

Date	N°	Nature	Title	Partners (resp.)
T0+50	ED4-9	Software	Final version multimedia platform	IPN, Cnam, LABRI
Cancel. (→ **C.6**)	ED4-10	Report	User evaluation report	Cnam (INA)

* In this project, INA is a subcontractor of Cnam.

E IMPACT OF THE PROJECT

E.1 IMPACT INDICATORS

Number of publications and communications (to detail in E.2)

		Multi-partner publications	Single partner publications
International	**Revues à comité de lecture**	2 published (+2 under submission)	
	Ouvrages ou chapitres d'ouvrage		
	Communications (conférence)	13 published/accepted (+2 submitted)	3
National (France or Mexico)	**Revues à comité de lecture**		
	Ouvrages ou chapitres d'ouvrage	1	
	Communications (conférence)		
Dissemination actions	**Articles vulgarisation**		2
	Conférences vulgarisation		
	Autres	4	

Autres valorisations scientifiques (à détailler en E.3)

	Nombre, années et commentaires (valorisations avérées ou probables)
Brevets internationaux obtenus	
Brevet internationaux en cours d'obtention	
Brevets nationaux obtenus	
Brevet nationaux en cours d'obtention	3
Licences d'exploitation (obtention / cession)	
Créations d'entreprises ou essaimage	
Nouveaux projets collaboratifs	« Gaiar » multimedia indexing platform – regional project in Aquitaine-Poitou-Charente-Limousin (in preparation).
Colloques scientifiques	
Others (specify)	MEXaction2 action detection and localization dataset http://mexculture.cnam.fr/xwiki/bin/view/Datasets/ 7 other datasets prepared

E.2 Liste des publications et communications

(the numbers allow to identify the publications on the web site of the project http://mexculture.cnam.fr/xwiki/bin/view/Main/Publications)

E.2.1 International: Journals with review committee

[11] Gabriel Sargent (CEDRIC & LABRI), Karina R. Perez-Daniel (LABRI & IPN), Andrei Stoian (CEDRIC & LABRI), Jenny Benois-Pineau (LABRI), Sofian Maabout (LABRI), Henri Nicolas (LABRI), Mariko Nakano-Miyatake (IPN), Jean Carrive (INA). A scalable summary generation method based on cross-modal consensus clustering and OLAP cube modeling. Under publication in Multimedia Tools and Applications (MTAP).

[12] Andrei Stoian (CEDRIC & LABRI), Marin Ferecatu (CEDRIC), Jenny Benois-Pineau (LABRI), Michel Crucianu (CEDRIC). Fast action localization in large scale video archives. Accepted for publication in IEEE Transactions on Circuits and Systems for Video Technology (IEEE TCSVT), DOI 10.1109/TCSVT.2015.2475835.

Submitted (April 2016): ACM Journal on Computing and Cultural Heritage (JOCCH).

Under preparation (for submission in June 2016): SPIE, JEI Special Section on Image Processing for Cultural Heritage.

E.2.2 International: communications

[1] Ballas N. et al. [including Andrei Stoian (CEDRIC & LaBRI), Jenny Benois-Pineau (LaBRI), Michel Crucianu (CEDRIC)]. IRIM at TRECVID 2012: Semantic Indexing and Instance Search. In Proc. of TREC Video Retrieval Evaluation workshop, 2012, 12p.

[2] Gabriel Sargeant (LaBRI), Pierre Hanna (LaBRI) and Henri Nicolas (LaBRI). Segmentation of music video streams in music pieces through audio-visual analysis. In Proc. of ICASSP, IEEE, Florence, Italy, May 2014.

[3] Karina Ruby Perez Daniel (IPN & LaBRI), Jenny Benois-Pineau (LaBRI), Sofian Maabout (LaBRI), Gabriel Sargent (LaBRI) and Mariko Nakano (IPN). Scalable Video Summarization of Cultural Video Documents in Cross-Media Space based on Data Cube Approach. 12th International Workshop on Content-Based Multimedia Indexing (CBMI 2014), Klagenfurt, Austria, June 2014.

[4] Atoany Fierro-Radilla (IPN), Karina Perez-Daniel (IPN), Mariko Nakano-Miyatake (IPN), Jenny Benois (LaBRI). Dominant Color Correlogram Descriptor for Content-Based Image Retrieval. 3rd Intl. Conf. on Image, Vision and Computing (ICIVC 2014), Paris, France, Sept. 2014.

[5] Montiel Pérez J. Yaljá (IPN), Torres Patiño Juan C. (IPN), Romero Herrera R. (IPN), Ramírez Acosta A. (IPN). Algoritmo para la extracción de frames representativos de video digital. Congreso Intl. en Ingeniería Electrónica. Mem. Electro 2014, Vol.36, pp. 250-254. Chihuahua, México. Oct. 2014.

[6] De la O Torres Saúl. (IPN), Martínez Nuño J. Alfredo. (IPN), García Vázquez Mireya S. (IPN), Hernández García Rosaura (IPN). 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, México. October 2014.

[7] Andrei Stoian (CEDRIC & LABRI), Marin Ferecatu (CEDRIC), Jenny Benois-Pineau (LABRI), Michel Crucianu (CEDRIC). Fast cascaded action localization in video using frame alignment. Intl. Workshop on Comput. Intelligence for Multimedia Understanding, 1-2 Nov. 2014, Paris, France.

[8] Atoany Fierro-Radilla (IPN), Karina Perez-Daniel (IPN), Mariko Nakano-Miyatake (IPN), Hector Perez-Meana (IPN), and Jenny Benois-Pineau (LABRI). An Effective Visual Descriptor Based on Color and Shape Features for Image Retrieval. Proceedings of 13th Mexican Intl. Conf. on Artificial Intelligence, MICAI 2014, Tuxtla Gutiérrez, Mexico, November 16–22, 2014, Part I, pp. 336-348.

[9] Andrei Stoian (CEDRIC & LABRI), Marin Ferecatu (CEDRIC), Jenny Benois-Pineau (LABRI), Michel Crucianu (CEDRIC). Scalable action localization with kernel-space hashing, IEEE International Conference on Image Processing (ICIP), Québec, Canada, 27-30 septembre 2015.

[10] Alejandro Ramirez (IPN), Jenny Benois-Pineau (LABRI), Mireya Saraí García Vázquez (IPN), Andrei Stoian (CEDRIC), Michel Crucianu (CEDRIC), Mariko Nakano (IPN), Francisco Garcia

Ugalde (UNAM), Jean-Luc Rouas (LABRI), Henri Nicolas (LABRI), Jean Carrive (INA). The Mex-Culture multimedia platform for the preservation and dissemination of the Mexican Culture. Content-Based Multimedia Indexing (CBMI), Prague, 10-12 juin 2015, demo.

[13] Atoany Fierro-Radilla (IPN), Karina Perez-Daniel (IPN), Mariko Nakano-Miyatake (IPN), Hector Perez-Meana (IPN), and Jenny Benois-Pineau (LABRI). An Effective Visual Descriptor Based on Color and Shape Features for Image Retrieval. Proceedings of 13th Mexican International Conference on Artificial Intelligence, MICAI 2014, Tuxtla Gutiérrez, MX, Nov. 16–22, 2014, Part I, pp. 336-348.

[14] Lester A. Oropesa Morales (IPN), Abraham Montoya Obeso (IPN), Rosaura Hernández García (IPN), Sara I. Cocolán Almeda (IPN), Mireya S. García Vázquez (IPN), Jenny Benois-Pineau (LABRI), Luis M. Zamudio Fuentes (IPN), Jesús A. Martinez Nuño (IPN), Alejandro A. Ramírez Acosta (IPN). Video annotations of Mexican nature in a collaborative environment. In Proc. SPIE Vol.9598 Optics and Photonics for Information Proc. IX. SPIE, 9598-24. San Diego, California, USA. 9–13 Aug. 2015.

[15] Abraham Montoya Obeso (IPN), Lester A. Oropesa Morales (IPN), Luis Fernando Váquez (IPN), Sara I. Cocolán Almeda (IPN), Andrei Stoian (CEDRIC & LABRI), Mireya S. García Vázquez (IPN), Luis M. Zamudio Fuentes (IPN), Jesús Y. Montiel Pérez (IPN), Saúl A. de La O Torres (IPN), Alejandro A. Ramirez Acosta (IPN). Annotations of Mexican bullfighting videos for semantic index. Part of Proceedings of SPIE Vol.9598 Optics and Photonics for Information Processing IX. SPIE, 9598-28. San Diego, California, USA. 9–13 August 2015.

[16] Sargent G. (LABRI & CEDRIC), Hanna P. (LABRI), Nicolas H. (LABRI) and Bimbot F. (LABRI). Exploring the complementarity of audio-visual structural regularities for the classification of videos into TV-program collections. In Proc. of IEEE International Symposium on Multimedia. Miami, Florida, December 14-16, 2015.

Accepted:

Proc. SPIE 2016 Optics and Photonics for Information. San Diego, California, USA.Aug. 2016, "New Generation of the Multimedia Search engines", IPN-LaBRI.

Proc. SPIE 2016 Optics and Photonics for Information. San Diego, California, USA.Aug. 2016, "Image Annotation for Mexican buildings database", IPN-LaBRI.

Submitted: ACM Multimedia 2016, Amsterdam, Netherlands, October 2016, IPN-LaBRI.

E.2.3 Book chapter (national)

[17] Jesús Montiel, Mireya García, Jenny Benoit, Michel Crucianu, "Plataforma Multimedia MEX-CULTURE", Collection: "Archivos Digitales Sustentables. Conservación y acceso a las colecciones sonoras y audiovisuales para las sociedades del futuro" Editor: Instituto de Investigaciones Bibliotecológicas y de la Información de la Universidad Nacional Autónoma de México, 2016.

E.2.4 Articles de vulgarisation

[18] IPN y Francia crean software, Gaceta del IPN, abril 2015. http://www.repositoriodigital.ipn.mx/bitstream/handle/123456789/21148/COM-085-2015.pdf?sequence=1

[19] Crean biblioteca digital que engloba 100 mil horas de audio, video e imágenes de la cultura mexicana, febrero 2016. http://www.cudi.edu.mx/noticia/crean-biblioteca-digital-que-engloba-100-mil-horas-de-audio-video-e-im%C3%A1genes-de-la-cultura

E.2.5 Dissemination actions

Demonstration of the Mex-Culture platform at "Futur en Seine" 2015. This included a video showcasing the project (now also available on the home page of the project http://mexculture.cnam.fr) and an interactive video summarization demo.

Presentation of the project and Mex-Culture platform at "Sustainable digital files" international congress, Research institute librarianship and information 2015. (http://iibi.unam.mx/micrositio/CIADS/programa.html).

To come:

1. Demonstration of the Mex-Culture results to Embassy of France in Mexico, Conacyt and Tijuana students community at "Dissemination seminar in Tijuana 2016" (may 27th). This included a video demo.
2. Mex-Culture project group at "Workshop in multimedia platforms applied to the preservation of the audiovisual cultural heritage as a priority theme of digital society" in Tijuana. This event is support by Embassy of France in Mexico, Conacyt and IPN (sept 5th and 6th 2016).

E.3 Liste des elements de valorisation

E.3.1 New collaborative projects

The "Gaiar" multimedia indexing platform is a regional project being set up in Aquitaine-Poitou-Charente-Limousin (France).
Digging into data challenge, Mexico is invited to participate in international project with USA and Netherlands (http://diggingintodata.org/), to submit in June 2016.

E.3.2 Open evaluation datasets

The MEXaction2 action detection and localization dataset developed in the project is publicly available on http://mexculture.cnam.fr/xwiki/bin/view/Datasets/Mex+action+dataset. It was referenced as a publicly available resource by CVOnline (http://homepages.inf.ed.ac.uk/rbf/CVonline/Imagedbase.htm#action) and also by Computer Vision Online (http://www.computervisiononline.com/dataset/mexaction2). MEXaction2 was presented to the organizers of TRECVID evaluation campaigns and to several research teams (including INRIA Willow team, University of Barcelona, University of Edinburgh).

2 national copyright by IPN (03-2015-111810003100-01, 03-2015-111810183000-01).
5 national copyright submitted by IPN in April 2016.

E.4 Bilan et suivi des personnels recrutes en CDD (hors stagiaires)

Identification				Avant le recrutement sur le projet			Recrutement sur le projet				Après le projet				
Nom et prénom	Sexe H/F	Adresse email (1)	Date des dernières nouvelles	Dernier diplôme obtenu au moment du recrutement	Lieu d'études (France, UE, hors UE)	Expérience prof. Antérieure, y compris post-docs (ans)	Partenaire ayant embauché la personne	Poste dans le projet (2)	Durée missions (mois) (3)	Date de fin de mission sur le projet	Devenir professionnel (4)	Type d'employeur (5)	Type d'emploi (6)	Lien au projet ANR (7)	Valorisation expérience (8)
STOIAN Andrei	H	andrei.stoian@gmail.com	15/12/2015	Master 2	France	-	Cnam	doctorant	36	30/11/2015	CDI	Grande entreprise	Ingénieur de recherche	Non	Oui
SARGENT Gabriel	H	gsargent@free.fr	18/12/2015	Doctorat	France	Doctorat	LaBRI	Post-doc	9	30/09/2014	CDD	Cnam	Post-doctorant	Oui	Oui
SARGENT Gabriel	H	gabriel.sargent@yahoo.fr	18/12/2015	Doctorat	France	Post-doc	Cnam	Post-doc	6	31/03/2015	CDD	Recherche publique	Ingénieur de recherche	Non	Oui

Aide pour le remplissage

(1) Adresse email : indiquer une adresse email la plus pérenne possible

(2) Poste dans le projet : post-doc, doctorant, ingénieur ou niveau ingénieur, technicien, vacataire, autre (préciser)

(3) Durée missions : indiquer en mois la durée totale des missions (y compris celles non financées par l'ANR) effectuées sur le projet

(4) Devenir professionnel : CDI, CDD, chef d'entreprise, encore sur le projet, post-doc France, post-doc étranger, étudiant, recherche d'emploi, sans nouvelles

(5) Type d'employeur : enseignement et recherche publique, EPIC de recherche, grande entreprise, PME/TPE, création d'entreprise, autre public, autre privé, libéral, autre (préciser)

(6) Type d'emploi : ingénieur, chercheur, enseignant-chercheur, cadre, technicien, autre (préciser)

(7) Lien au projet ANR : préciser si l'employeur est ou non un partenaire du projet

(8) Valorisation expérience : préciser si le poste occupé valorise l'expérience acquise pendant le projet.