THE MAVEN PROJECT: MANAGEMENT AND AUTHENTICITY VERIFICATION OF MULTIMEDIA CONTENTS

C. Ruiz[*], *I. Krsteski*[†], *P. Dago*[‡], *J. Sánchez*[‡], *L. Pérez-Freire*[‡], *A. De Rosa*[″], *D. Ariu*[‴], *L. Piras*[‴], *R. Ahumada*[ǁ], *M. Jerian*[❦]

[*]Taiger, [†]Arthaus, [‡]GRADIANT, [″]CNIT, [‴]UNICA, [ǁ]XTREAM, [❦]Amped
carlos.ruiz@taiger.com, igork@arthaus.mk, {pdago, jsanchez, lpfreire}@gradiant.org, alessia.derosa@unifi.it, davide.ariu@diee.unica.it, luca.piras@diee.unica.it, rahumada@xtreamsig.com, martino.jerian@ampedsoftware.com

ABSTRACT

MAVEN (Management and Authenticity Verification of multimedia contENts) is a European FP7 Project focused on the development of a suite of tools for multimedia data management and security. MAVEN objectives are centered on two key concepts, search and verify, integrated in a coherent manner: the system first searches for digital contents containing objects of interest and then applies advanced forensic analysis tools to verify their integrity and authenticity. These capabilities have been developed as a single software framework, and the project also involves the implementation of a prototype demonstrator application, which brings to the end user the possibility of searching for specific contents in media while *verifying their authenticity.*

Index Terms— media analysis, multimedia search, forensics, EU project

1. INTRODUCTION

The 21st century society is universally recognized as the information and communication society. Information is continuously generated, acquired, and shared. A large part of this information is stored within multimedia documents generated in a number of different scenarios. It must be also considered that the availability of low-cost, high-capacity storage devices makes easy to quickly accumulate thousands of multimedia files.

Thus, the efficient management of large amounts of multimedia files is indeed a challenging task. In addition, it is common knowledge that digital assets are extremely volatile, in the sense that digital documents can be easily edited, intentionally or unintentionally, so that their content can be modified and the conveyed information can change its meaning. It is incontrovertibly true that digital documents are natively more prone than others to modifications and tampering: thus, in order to make this information valuable, it is fundamental to verify the integrity of the document for assuring the authenticity of the associated information. Governments, national and international associations are aware of the fact that the phenomena may also have legal, ethical, social, and cultural implications.

The MAVEN Project[1] addresses these issues by using some of the latest technologies, powering integrity and authenticity verification tools with multimedia analysis algorithms that search for specific contents. In particular, the MAVEN capabilities range from face detection and recognition to image source verification. All the different modules are integrated within the same framework for application development.

2. THE CONSORTIUM

The consortium behind the MAVEN project is formed by a group of four SMEs involved in business areas directly related to the search and verification of multimedia contents: AMPED (Italy), ARTHAUS (Macedonia), TAIGER [2] (Austria and Spain) and XTREAM (Spain). The MAVEN consortium also comprises three RTD performers with complementary expertise and a strong background in the technological areas related to MAVEN: CNIT (Universities of Siena and Florence; Italy), the Pattern Recognition and Applications group from the University of Cagliari (Italy) and GRADIANT (R&D Center, Spain).

3. THE MAVEN PROJECT

Despite the technological advances in the Security and Media sectors, MAVEN arises from the need of providing such industries with a suite of advanced solutions able to operate in a range of realistic scenarios (CCTV, web images, broadcast data, etc). Moreover, the search and verify concept is not supported in an integrated manner by any tool currently available in the market. The MAVEN suite combines state-of-the-art multimedia analysis techniques with verification algorithms for assessing the authenticity of media assets, providing significant benefits:

- Comprehensiveness: MAVEN integrates seven

[1] http://www.maven-project.eu/
[2] Formerly known as playence

different technologies into a unique toolbox, providing both efficient search capabilities and solutions for guaranteeing the authenticity and integrity of digital contents. Table 1 describes the tools developed in the project.

- Robustness: In contrast to current solutions, the MAVEN suite has been designed to provide robustness in realistic scenarios by using state-of-the-art algorithms and methodologies.

Table 1. The different tools in the MAVEN framework

Name	Description
Image Source Identification	Tool for linking a digital photo to its corresponding camera device, in order to identify image origin.
Image Integrity Verification	Tool for verifying if a digital image suffered some local/global editing or manipulation (with original image available/not available).
Video Integrity Verification	Tool for assessing if a digital video suffered double encoding and insertion/ removal of frames.
Text Localization and Extraction	Tool for detection and recognition of text within unconstrained scene imagery.
Spoken Keyword Detection	Tool for the spotting of (previously known) keywords in audio tracks.
Face Detection and Recognition	Tool for the automatic detection and recognition of faces in image and video galleries.
Object and Scene Recognition	Tool for detection of a set of company logos and categorization of scenes.

4. FORENSIC ANALYSIS

The project has dealt with the problem of integrity and authenticity verification of digital visual contents, i.e. images and video sequences; a set of modules has been developed mainly based on multimedia forensics technologies, in order to answer two main questions:

i) What is the source device for a digital image?

ii) Has the digital object undergone some editing process?

In particular, the project focused on four different use cases, each with specific requirements. According to such needs, a proper set of technologies has been developed.

Image Source Identification - A user wants to link his/her digital photos to his/her corresponding camera devices. To this aim, watermarking technology is applied in order to embed within the digital content some information (i.e. the digital watermark) describing the camera device; at a later time, it possible to check whether an image found on the web contains the watermark matching a specific camera information.

Informed Image Integrity Verification - A user wants to verify if a set of retouching operations has been applied to an image, by comparing such an image with the corresponding original one. In order to make easier this comparison and having the original images available, some dissimilarity metrics between the processed and the original image have been applied, based on the discipline called Change Detection.

Blind Image Integrity Verification - A user wants to understand if a digital image is authentic or suffered some manipulation with photo editing software. In this case original images are not available, so image forensics technologies, that are able to capture important information on the image history without any a-priori knowledge, are exploited. Since the analyst cannot know in advance which is the most appropriate tool to be used on a suspect image, a set of different and complementary forensic tools has been developed and their outputs intelligently merged.

Video Integrity Verification - A user wants to understand if a video has been recompressed or not, as well as if some frames have been removed or added in the sequence. Also in this case we have access only to the to-be-analyzed video, thus exploiting video forensics technologies. In particular, by starting from a video double encoding detection algorithm, we designed a module that localizes whether a misalignment in the frame structure of the video occurred between two successive encodings.

5. OBJECTS AND SCENE ANALYSIS

The Objects and Scene Analysis component actually consists of one tool responsible for providing text localization and extraction functionalities, whereas the second is aimed to provide the capability to detect a particular content within image and video galleries. More in details, the tools allows to categorize a scene (e.g. to say if the scene represents a kitchen, a garden, a sky), to detect the presence of a particular kind of object (e.g. a sofa, a bed), and also to detect the presence of a particular company logo.

The text localization and extraction module consists of two sub modules, one focused on the localization and the second focused on the extraction of the text. Both text detection and recognition functionalities have been implemented through state-of-the-art techniques available in the literature. The devised text detection approach has been oriented to handle the specific case of "superimposed text" on video frames, in order to take advantage of the specific constraints of this kind of text (such as colour, size, position, and above its invariance in multiple consecutive frames). In particular, a mixed methodology using Class Specific Extremal Regions (CSER) and a perceptual organisation framework has been implemented, in the fashion of [Neuman2012]. The text detection module has been adapted to the target scenarios and has been validated with the ICDAR 2013 scene text dataset.

The text recognition sub-module exploits a standard OCR engine (Tesseract) for the recognition the word's binary image, generated through a proper pre-processing phase. The sub-module is able to receive from the text localization sub-module text areas which might correspond either to a single line of words of the same size or, in some cases, to multiple lines of words. Before being analysed by the OCR, the text area is first normalized in size and binarized. Before

being processed by the OCR, the image is further cleaned to remove noisy artefacts and other elements that can affect the subsequent character recognition, like some horizontal lines near the characters. The module has been evaluated on the Challenge 2 (Reading Text in Scene Images) of the ICDAR 2013 Robust Reading Competition, and exhibited a Word Recognition rate of 57% (fast-mode) and 62% respectively), the ICDAR 2013 baseline being 45.3%

The Object and Scene recognition module tool consists of two sub-modules. The one responsible for Logo Recognition implements an algorithm based on the match between keypoints of two images. Keypoints are usually corners, edges or points characterized by an abrupt variation of luminosity in one or more directions. We used DNets [Hundelshausen, 2012] as descriptors, which were formerly designed to evaluate similarity between full images and not between portions of them, and adapted their functioning to the case of logo detection and recognition. The module, evaluated on the Flickr32 Logo Dataset, exhibited a Recall value (54%) comparable with that of the state of the art algorithms (61%).

The object and scene sub-module implements instead a state of the art approach based on the HOG descriptors, where every image is processed and a histogram of oriented edges (HOG) descriptors is extracted on a regular grid at steps of 8 pixels. The HOG features are extracted for all the images in the training set and the resulting descriptors are then clustered in 300 visual words by a k-means algorithm, to build a set of visual words. Such visual words are used to compute a histogram of the HOG descriptors, in a bag of word (BOW) way, to represent the image. Classification of object and scenes is based on a SVM classifier, trained in a one-VS-all configuration. Performance depends on the number of categories considered, decreasing with them. Randomly sampling 15 out of the 397 categories of the SUN11 dataset, we obtained an average precision of 68%, which decreases to 61% if the number of classes is raised to 25. In the case of objects, the average precision measured on the 20 classes of the PASCAL VOC dataset 2012 is 49%.

6. HUMAN-TRAIT ANALYSIS

Two of the tools developed in the MAVEN project are devoted to the analysis of human traits, namely the spoken keyword detection module and the face detection and recognition module.

The spoken keyword detection tool is based on Hidden Markov Models (HMM). It has been tested on TIMIT and Albayzin databases and showed a performance comparable or better than state of the art systems. It also features an enhanced scoring algorithm than increases the accuracy on short words. In addition, it is able to search for a keyword several times faster than real time.

The face detection and recognition tool included the development of face detection and facial recognition modules. The face detection module was evaluated following the FDDB benchmark [Vidit2010], and outperformed the Viola&Jones detector implementation in OpenCV [3] obtaining 70% of True Positive Rate (TPR; working point was set on 250 false positives), while Viola&Jones detector obtained a TPR of 45% (also at the working point of 250 false positives). In addition, real time processing has been achieved, allowing face detection faster than real time in high-resolution videos.

The face recognition module is responsible for assigning a face an identity picked up from a database of known ones, thus complementing the face detection module described above. The implemented algorithm is based on the BSIF (Binarized Statistical Image Features) [Kannala, 2012]. The module has been evaluated on the Experiment 4 of Face Recognition Grand Challenge Database, where 80% of Face Verification Rate has been achieved at False Acceptance Rate 10%. The module is able to deal with small rotations (+/- 10º in-plane and +/- 30º out-of-plane) and with small occlusions. Experiments have been carried out within uncontrolled lighting conditions and considering faces where the minimum distance of 64 pixels exists among the eyes.

7. THE MAVEN TOOLKIT

The MAVEN project aims to develop a suite of tools for multimedia data processing, by including and combining different functionalities and capabilities. The implementation of these tools has been carried out in C++, and its final form will be a set of software libraries implementing a specific API. This collection of libraries will thus serve as a Software Development Kit (SDK) for building new commercial applications. The initial approach to the API design follows certain principles, which include the modularity and decoupling of packages implementing different functionalities, the abstraction towards the low-level implementation of the modules, the maintainability and understandability of the API, and the user friendliness of MAVEN as a software product. These principles are commonly followed in large software projects as they ensure a correct development of the product and ease its maintenance, therefore extending its lifespan.

In order to follow the aforementioned principles, a multi-layer scheme has been followed towards the structural design of the API. The bottom layer of this structure (development API) is formed by the different modules (i.e. forensic tools and search tools) developed by the corresponding RTDs. The mid-level layer, also referred as the core API, will encapsulate these modules into well-defined libraries, defining the necessary data structures and interfaces for the joint management of the different tools. A third and top level API (application service API) is incorporated with the aim of providing an additional level of abstraction towards the demonstrator application and to future applications. From the end user's point of view, only the two higher layers of the API will be visible, since the

[3] http://opencv.org/

bottom layer will be wrapped by the mid-level interfaces. Figure 1 shows the structure of the SDK.

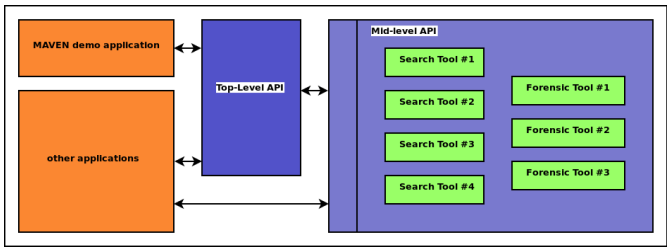

Figure 1: MAVEN toolkit structure

8. EXPECTED IMPACT

MAVEN has the potential to significantly change the way that many companies and organizations manage their multimedia assets. The results of MAVEN will allow SMEs, law enforcement bodies, press agencies, insurance companies, and broadcasting companies, among others, to manage their multimedia contents and verify their integrity and authenticity in an efficient and scalable manner. The SMEs participating in the project will play an important role in maximizing the impact of the developed technologies, as explained below.

AMPED will disseminate the results mainly to law enforcement labs and government agencies, which will benefit from MAVEN advanced tools for forensics and intelligence activities. In particular, the verification of video integrity is expected to gain wide acceptance, since video authentication is a very difficult topic without tools currently available in the market.

ARTHAUS will disseminate the results in professional communities of web application developers dealing with large image databases. The first use case to be tested will be in an application for capturing, processing and delivery of professional, high quality, real-estate images. Such an application faces the problems of resolving image ownership, image cataloguing, and in particular conformance checking of image retouching. The MAVEN tools successfully address all those problems.

TAIGER will explore the application of MAVEN technologies in new markets where the technology has the potential to open new opportunities. In a first step, MAVEN will be integrated with products already available at TAIGER (in particular, the product iSearch[4]). In a second step, TAIGER will explore the exploitation of MAVEN under a SaaS model, which has the potential to reach massive audiences bringing the benefits of MAVEN tools to virtually all organizations, in particular small companies.

Finally, XTREAM will focus on the dissemination and exploitation of MAVEN technologies in market segments where it already possesses a well-developed distribution and partnership network, in particular Government and Homeland Security, where MAVEN addresses real market demands. Such dissemination will be facilitated in a first step by integrating MAVEN modules in two products of XTREAM portfolio.

[4] http://www.taiger.com/isearch/

9. CONCLUSIONS

MAVEN is a European FP7 Project focused on the development of a suite of tools for multimedia data management and security. MAVEN project integrates state-of-the-art media analysis algorithms to analyze and verify multimedia content. These algorithms have been developed within MAVEN and have been evaluated using both synthetic and real-world data sets.

Moreover, the industrial partners of MAVEN, which are leaders in their fields, will ensure the wider industrial uptake of the technologies developed in the project, by adopting them in their products and by developing innovative business models that take advantage of them. The commercial impact of MAVEN includes short term and medium-to-long term results by following appropriate individual dissemination, exploitation mechanisms and business plans (e.g. integrating MAVEN technologies to pre-existing successful systems, products and services).

10. REFERENCES

[Kannala2012] J. Kannala. J and E. Rahtu. "BSIF: Binarized Statistical Image Features". In *Proc. 21st International Conference on Pattern Recognition (ICPR 2012)*, pp 1363–1366, 2012.

[Hundelshausen2012] F. Von Hundelshausen and R. Sukthankar, *D-Nets: Beyond Patch-Based Image Descriptors*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012

[Neumann2012] L. Neumann, and J. Matas, (2012). *Real-time scene text localization and recognition*. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 3538-3545.

[Vidit2010] Vidit Jain and Erik Learned-Miller. FDDB: A Benchmark for Face Detection in Unconstrained Settings. Technical Report UM-CS-2010-009, Dept. of Computer Science, University of Massachusetts, Amherst. 2010.