Toward an Integrated System for Surveillance and Behaviour Analysis of Groups and People

Edoardo Ardizzone, Alessandro Bruno, Roberto Gallea,
Marco La Cascia, and Giuseppe Mazzola

Dipartimento di Ingegneria Chimica, Gestionale, Informatica, Meccanica, Università degli Studi di Palermo, Palermo, Italy
{edoardo.ardizzone,alessandro.bruno15,roberto.gallea,
marco.lacascia,giuseppe.mazzola}@unipa.it

Abstract. Security and INTelligence SYStem is an Italian research project which aims to create an integrated system for the analysis of multi-modal data sources (text, images, video, audio), to assist operators in homeland security applications. Within this project the Scientific Research Unit of the University of Palermo is responsible of the image and video analysis activity. The SRU of Palermo developed a web service based architecture that provides image and video analysis capabilities to the integrated analysis system. The developed architecture uses both state of the art techniques, adapted to cope with the particular problem at hand, and new algorithms to provide the following services: image cropping, image forgery detection, face and people detection, weapon detection and classification, and terrorist logo recognition. In the last phase of the project we plan to include in our system new services, mainly oriented to the video analysis, to study and understand the behaviour of individuals, either alone or in a group.

Keywords: homeland security, weapon detection, weapon classification, image analysis, video analysis, logo recognition, forgery detection, information fusion.

1 The SINTESYS Project

SINTESYS (Security and INTelligence SYSstem) [1] is an Italian research project, funded by the PON R&C 2007-2013 program, which involves many scientific (Università del Salento, Università degli Studi di Palermo, Università degli Studi di Salerno, ICAR-CNR, CeRICT) and industrial (Engineering Ingegneria Informatica, Expert System, Digital Video, System Management) partners. The goal of SINTESYS (Security and INTelligence SYSstem) is to study, define and develop new technologies for the realization of an innovative integrated system which can analyse, plan, investigate 'open' multi-modal (text, images, video, audio, ...) data sources (OSINT - Open Source INTelligence) in an integrated, coherent and consistent way, in order to discover the presence of links, relationships, connections which the disjointed evaluation of individual sources would not be able to highlight, and thus to give an important contribution to the management of the case of interest in terms of Decision Support System. SINTESYS is

A. Petrosino, L. Maddalena, P. Pala (Eds.): ICIAP 2013 Workshops, LNCS 8158, pp. 474–481, 2013.

mainly, but not only, directed to analysts who work in institutional sectors, supporting them in the most advanced Intelligence processes for homeland security applications, through collection, processing, analysis and distribution of information. For this purpose, SINTESYS uses and combines techniques, technologies and innovative models for sound analysis, image recognition, movie recognition, social network analysis, text mining, human computer interaction, cognitive psychology, along with models and techniques for information fusion and artificial intelligence. SINTESYS is also establishing new models and innovative techniques to analyse the social dynamics within groups and communities, in order to provide hidden information which may be of importance for security issues. SINTESYS therefore proposes to create an integrated software system equipped with advanced tools for analyzing and correlating vast amounts of data from a variety of heterogeneous, multichannel and multimodal information sources. Using innovative techniques of feature extraction in a combined and integrated way on the same contents, enables synergy and "disambiguation" of data. This analysis, together with a study on correlation of the same data, leads to the emergence of situations of potential danger to public security, ranging from recognition and localization of socially dangerous groups of people or individuals, to the discovery of communicative dynamics which may suggest the need to monitor for prevention aims. With this goal the SINTESYS partners defined a taxonomy of interesting macro-events, which spans from People Detection, to Behavioural Recognition, to Terrorist Attacks, that the system should be able to detect, by analyzing and fusing micro-events, that are detected by each separate analysis module. Furthermore, the system is designed to be flexible to suit the various needs of intelligence analysts, who are able to navigate data using intelligent graphical user interfaces which adapts to different types of information sources and to the actual user needs, through a specific survey of the interaction habits and a psychological study on interaction patterns. The project is still in progress, and currently both state of the art technologies and new algorithms, with particular attention to open source environments, have been developed by the involved research units. We expect that the innovative character of the project results will lead to the growth of various markets and sectors related to the homeland security, individual behaviour understanding and social relationship analysis.

2 Image and Video Analysis

Within the SINTESYS project the Scientific Research Unit (SRU) of University of Palermo) is responsible for image and video analysis. For this purpose the SRU in the first year of the project, studied and developed algorithms for medium level feature extraction, for image data sources: a saliency based image cropping service, a state of the art face detection and people detection algorithms, two proposed methods for the detection and the classification of weapons, a service for the recognition of logo of terrorist groups, and an image forensics technique to verify the authenticity of a digital image. These algorithms are made available as web services and accessed by the integrated analysis system. In the next subsections we will describe the client-server architecture of the image and video analysis system, and the implemented services.

2.1 System Architecture

The selected architecture is based on the client–server model. The structure of the web services is distributed between client (service requester) and the server (service provider). The client and the server communicate over http interface.

The client-server communication is handled with a computer network. The request of the *client* is sent through http interface as it follows:

```
http://sintesys.dicgim.unipa.it/sintesys.php?service =
service_name&url=image_url&par=[par1,par2,…,parN];
```

The client request is made of the web server address (*http://sintesys.dicgim.unipa.it/*), the name of the php file that handles the executions of the services (*sintesys.php*), the name of the service (*service_name*), the URL of the input image (*image_url*) and a vector of parameters, if needed. After the execution of the web service by the server, the results are displayed via the http interface using a JSON object. A JSON Object is an unordered collection of name/value pairs. As an example:

```
{"name":"http://sintesys.dicgim.unipa.it/out/image_crop/c
at_shot_by_arrow-1_out.jpg","bbox":[133,114,918,865]}
```

The output JSON Object includes the URL of the processed image, if any, that is temporarily stored into the server and the other results of the invoked services (the bounding box in this case, for the image cropping service).

The platform of the SINTESYS project is distributed across multiple *servers*. Our SRU, which deals with image processing, installed the developed applications onto a web server that handles the processing of the data via Apache Web Server and the Matlab Application Server. Once the client request is sent via HTTP interface, the Apache server interprets the request of the service by analyzing the name of the service. More particularly, a PHP application handles the selection of the requested web service by extracting the name of the service, the input image and any parameters from HTTP interface, as described in the previous section. The PHP application also handles the execution of the service through the Microsoft Component Object Model (COM) protocol that allows the selection of the specific application, related to the requested service. In our project the services are implemented in Matlab code, but the source code can be written in any of the many programming languages that support COM. Upgrades to applications are simplified, as components can simply be swapped without the need to recompile the entire application. In addition, a component's location is transparent to the application, so components can be relocated to a separate process or even a remote system without having to modify the application. Matlab function returns a JSON Object, as described in the above section, which have different fields, with respect to the different web services.

2.2 Services

Within the SYNTESIS project we studied and developed several algorithms for medium level feature extraction and to support the analysis of an inspected image source.

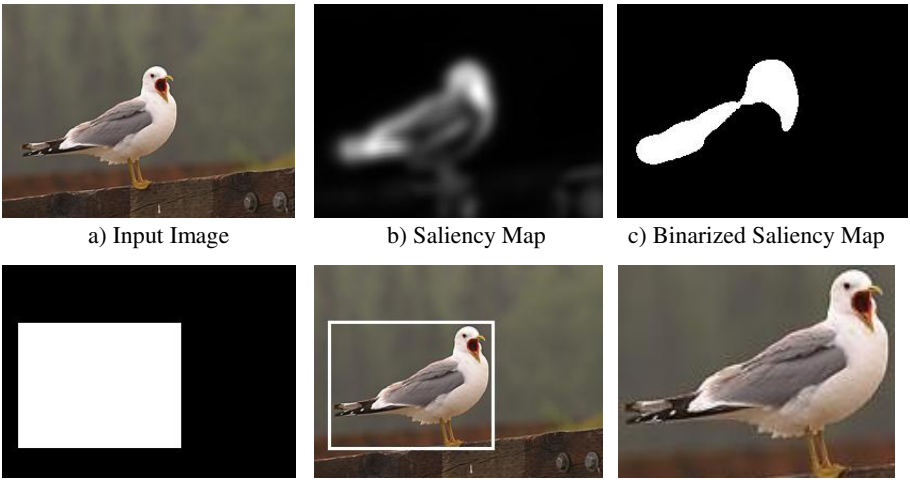

Fig. 1. The steps of the Image Cropping Method pipeline. a) Input Image, b) Saliency Map, c) Binarization, d-e) Bounding Box, e) Cropping.

2.2.1 Image Cropping

Image cropping is a technique that is used to resize an image by selecting its most relevant areas, discarding its useless or redundant parts. Our SRU developed a saliency based image cropping technique, which extracts salient information from the image to select the image crop. Our system can be subdivided into (see fig.1): Saliency Map Extraction, Saliency Map Binarization (Thresholding), Bounding Box Extraction, Photo Cropping. In our work we used the GBVS algorithm[2], which is one of the most popular state-of-the-art technique to extract the saliency map. The saliency map is then binarized using a threshold, which is experimentally set, and then the bounding box of all the pixels, which values are above the threshold, is selected and used to crop the photo (fig.1.f).

2.2.2 Face Detection

The Face Detection service is used to detected one of more faces into an input image. The implemented algorithm is based on the well-known Viola Jones descriptors [3], which are now adopted as a standard for the detection of faces in digital images.

2.2.3 People Detection

The People Detection service is based on the Dollar et al. algorithm [4]. It is designed to detect the position of people standing in the scene. The implemented technique use the Histogram of Gradients (HoG) which analyzes the distribution of the gradient of the image along different directions. Fig. 2 shows a visual example of the obtained results.

Fig. 2. A visual example of the results obtained with the implemented People Detection technique

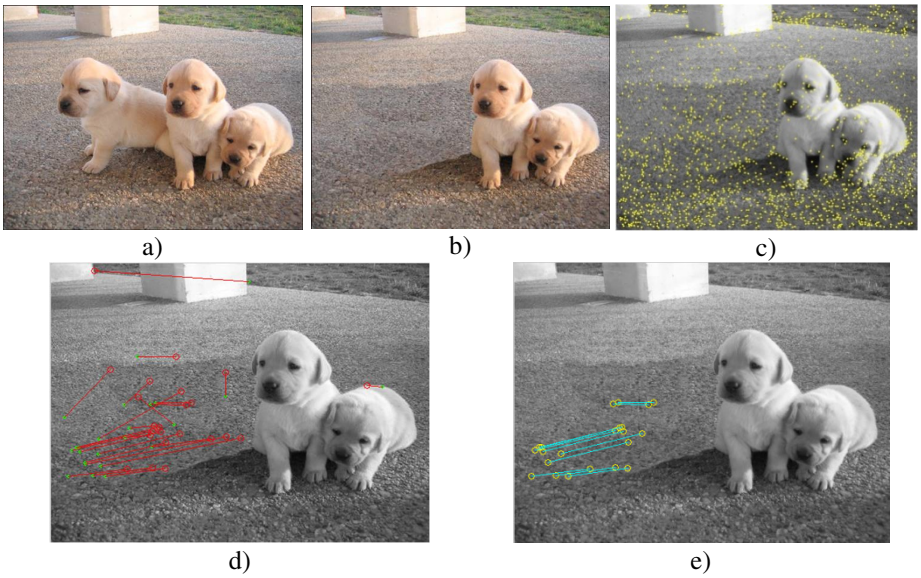

Fig. 3. Original image (a), Tampered image (b), superimposed SIFT keypoints (c), matching keypoints (d), after RANSAC (e)

2.2.4 Forgery Detection

The Forgery Detection service aims to verify the authenticity of a digital image, discovering if the image has been tampered by someone. The implemented algorithm is

based on the work of Pan and Lyu [5], and it is based on the extraction of the interest points. Points are then matched, comparing their local features, to identify possible duplicated regions. Matches are filtered by using the RANSAC algorithm, to find groups of points that match according to an affine transformation (see fig. 3). This method proved to be robust to geometric transformations (rotation and scaling), but cannot be used whenever the copy-paste region is uniform. In that case, block-matching methods are preferable, but they are extremely slow, therefore we decided to use a point-based technique.

2.2.5 Weapon Detection

The Weapon Detection service is a new method that is designed to detect whether an input image is a weapon or not. The proposed algorithm initially segments the image (see fig.4), by thresholding the saliency map, as discussed for the Image Cropping service, and refines the segmentation by using active contours [6]. It then extracts features from the foreground of the scene: texture (Edge Histogram [7]), color (Color Histogram) and shape (Turning Angle [8] Histogram). These descriptors are used to train three separate SVM classifiers, and the image is classified as weapon by majority. Weapon images are taken from the Internet Movie Firearms Databases [9]. Negative examples are images of objects that "can be hand-held" (books, bags, umbrellas, etc.) and have been downloaded by the Google Image Search Engine. Experimental results shows that our weapon detector achieves above the 90% of accuracy.

2.2.6 Weapon Classify

The Weapon Classification service is able to classify an image of a weapon into one of the following categories: gun, revolver, rifle, machine gun, heavy machine gun (fig. 5). In this case there is not a "no weapon" class, as the service supposes that the input image represents a firearm. The service extracts the HoG descriptor from the image, and uses a trained multiclass SVM to classify it. Training images are taken from [9]. The classification accuracy is, also in this case, above the 90%.

2.2.7 Recognition of Terrorist Logos

The Terrorist Logo Recognition service, compares an input image with a dataset of images representing the logos of 13 of the most known terrorist groups in the world. We built a reference dataset by downloading from the Web 10 different instances of each logo of the 13 selected classes (some examples in fig.6). The implemented algorithm is a KNN classifier, which compares the SIFT interest points and descriptors of the input image to the descriptors extracted from a reference dataset images. The algorithm proved to be robust to scaling, translation and rotation, as the SIFT descriptors are robust to affine transformation. The average accuracy of the method is about 65%.

Fig. 4. Weapon Detection: input image (a), segmentation mask (b)

Fig. 5. Some visual examples of the weapon classes

Fig. 6. Some visual examples of the logo classes


Fig. 7. Some case studies. Photos downloaded from the Web

3 Conclusions and Future Works

Even though computer vision techniques are not yet able to resolve the complete image understanding problem, experiments on real data showed that they can play a significant role in particular cases and constrained scenarios (see fig.7). Moreover when they are used in an integrated system they can take advantage from the analysis of other sources of information, or they can give hints to the other analysis modules. Each module is designed to find micro-events that, fused with the other source information, could help experts in the detection of the macro-events of interest. The future works of the Palermo SRU will focus on the video analysis techniques. In fact, the spread of the broadband in the world is increasing the availability of video resources on the Web, therefore they represent a very important source of information. In particular we are working on algorithms about people and object tracking, action recognition, video saliency detection, video summarization and skimming, and video forgery detection. The integration of the these video analysis services into the system will give an essential contribution toward the understanding of the human activities in social context, and of the interaction of individuals with other people or objects.

References

1. http://sintesys.eng.it/en_GB/home
2. Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. In: Advances in Neural Information Processing Systems 19, pp. 545–552. MIT Press (2007)
3. Viola, P., Jones, M.: Robust real-time face detection. International Journal of Computer Vision 5(2), 137–154 (2004)
4. Dollár, P., Belongie, S., Perona, P.: The Fastest Pedestrian Detector in the West. In: BMVC (2010)
5. Pan, X., Lyu, S.: Region duplication detection using image feature matching. IEEE Transactions on Information Forensics and Security, 857–867 (2010)
6. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. Int. J. Comput. Vision 1, 321–331 (1987)
7. Won, C.S., Park, D.K.: Efficient Use of MPEG-7 Edge Histogram Descriptor. ETRI Journal 24, 23–30 (2002)
8. Niblack, W., Yin, J.: A pseudo-distance measure for 2D shapes based on turning angle. In: Proceedings of the Int. Conference on Image Processing, vol. 3, pp. 352–355 (1995)
9. http://www.imfdb.org/wiki/Main_Page