

## Mobile Interface for Content-Based Image Management

Marco La Cascia, Marco Morana, Salvatore Sorce

DINFO – Dipartimento di Ingegneria Informatica  
Universita' degli Studi di Palermo  
Palermo, Italy

{lacascia|sorce}@unipa.it, morana@dinfo.unipa.it

**Abstract**—People make more and more use of digital image acquisition devices to capture screenshots of their everyday life. The growing number of personal pictures raise the problem of their classification. Some of the authors proposed an automatic technique for personal photo album management dealing with multiple aspects (i.e., people, time and background) in a homogenous way. In this paper we discuss a solution that allows mobile users to remotely access such technique by means of their mobile phones, almost from everywhere, in a pervasive fashion. This allows users to classify pictures they store on their devices. The whole solution is presented, with particular regard to the user interface implemented on the mobile phone, along with some experimental results.

**Keywords**—mobile interfaces; pervasive systems; CBIR - Content Based Image Retrieval

### I. INTRODUCTION

In the ubiquitous computing vision, technology is seamlessly integrated into the environment and provides useful services to humans in their everyday lives [19]. Real-life environments become *augmented* when they are spread with a large variety of smart devices, thus arranging a pervasive system.

In this vision, wireless connections play a relevant role, allowing designers to distribute the augmented environment intelligence among different remote devices, and allowing users to access distributed intelligence from everywhere.

The advent of small, personal mobile devices, and the new possibilities offered by third-generation wireless communication technologies, significantly contributed to the realization of such vision.

Mobile devices are widely diffuse among people of all ages; they are almost in everyone's pocket and can be used almost anywhere. This justifies the common interest in the study of new ways of service provision (and brand-new services too) according to the features and capabilities of mobile devices [20] [21].

There is a large variety of application fields where services can be pervasively accessed by mobile devices, such as context-aware information provision within university campuses [27], interactive user profile-based guides in cultural heritage sites [22] [24], augmented reality objects assembly in mobility [25], healthcare systems [23] [26].

Due to their increasing computing and memory capabilities, personal mobile devices are used for different purposes, such as instant messaging, multimedia storage and playing, and picture acquisition.

In particular, the embedding of small cameras within smartphones makes the users freedom to take pictures of everyday life with no need to carry them on different devices. The growing number of personal pictures taken at any time, along with the need to share them with friends and relatives, give rise to social networks such as Facebook, Flickr, and so on. Their millions of users demonstrate how large is the phenomenon.

In a previous work, we faced with the problem of classifying pictures according to some given criterion. In particular, we addressed the scenario in which a user takes pictures in different sessions and different places, that is pictures belong to different contexts. In this case users may also be interested in using automatic techniques to instantly manage (i.e., browse, save, print and so on) a subset of captured pictures according to some particular picture properties.

Currently, the main way to search digital photo libraries is by keywords given by the user. However this process has been observed to be inadequate since users add few, ambiguous, keywords for large set of images. This is even worse on mobile devices due to limited text input capabilities.

An ideal system for image browsing should allow an automatic organization of pictures based on the semantic of photos. Our point is that personal libraries stored on mobile devices show peculiar characteristics compared to general image collection. Namely the presence of people in most of the images and a relatively small number of different individuals across the whole library that allow to achieve reliable results with automatic approaches [1] [2].

In particular, in personal photo collection the user is mainly interested in *who* is in the picture (usually a relatively small number of different individuals) and *where* and *when* the picture was shot. Who, where and when are the fundamental aspects of photo information and input images can be intrinsically split in three domains of interest.

To automatically organize image data based on faces, background and time descriptors a mean-shift based approach is used [3].

In this paper we discuss an evolution of our system, in which users can manage their pictures on mobile phones.

Using a communication wrapper, we can remotely control the execution of our classification algorithm which runs as Matlab script on a remote machine by means of a suitable user interface running on the smartphone.

## II. RELATED WORKS

Due to the widespread diffusion of digital photography, several studies addressed the problem of semi-automatic personal photo collection management [4] [5] [6]. Most of the effort has been devoted to finding techniques to help the user in annotating the collection. For example iPhoto 09 [8], recently proposed by Apple, allows users to organize their libraries by using semi-automatic features detection (i.e., faces, places and events) and search them by person, location, title, album, event, or keyword.

Previous work regarding CBIR on mobile devices has been mainly limited to particular problems like the generation of an initial query set [10] or to energy efficiency [11]. In [13] a server application performs the CBIR and an internet browser is used on the mobile device to view the results. Some work focused on the development of JAVA-based client/server framework where mobile devices are used as front-end to traditional CBIR systems [12] [13]. However the authors simplistically consider mobile devices as standard devices that can be moved. Our point of view is quite different since nowadays mobile devices can be viewed as embedded systems that allow for ubiquitous computing.

To the best of our knowledge, there are few works concerning the use of a mobile device for some kind of image processing. Among them, in [9] authors introduced a real-time face processing application that make use of a suitable model to accomplish the portable face recognition task.

## III. SYSTEM ARCHITECTURE

### A. Overview

Matlab is a matrix-oriented computational environment, programmable by means of a script language which makes it highly dynamic and modular. There are many existing toolboxes by which programmers can suitably use Matlab for almost any purpose. One of the most used toolboxes is the IPT (Image Processing Toolbox), which enables users to execute almost all image transformations and operations, ranging from the simplest rotation and resizing, to the complex ones such as histogram equalization, Fourier transform computing, noise reduction filtering, morphologic operators application, and so on. Users can also write their own scripts to implement further operators. Such features make Matlab a precious tool for image processing.

The main drawback is that people that want to write their own algorithm have to be particularly skilled in Matlab programming in order to take advantage of all its features to carry on their own computations. On the other hand, even if

people want to use an already Matlab-coded algorithm, they must know how to deploy it, or where it is available to be fed with their own inputs.

The solution we discuss in this paper allows common people to run our algorithm on personal images, by means of personal mobile devices, such as Smartphones, provided that they are Java-enabled. In more detail, we make Matlab scripts, and in particular our image classification algorithm, available on a pervasive infrastructure, thus allowing users to use them with no need to know where they are actually running. Furthermore, we implemented a suitable user interface on the smartphone side which make the human-computer interaction more simple and intuitive. Exploiting the features of recent devices, such as accelerometers, the call for the algorithm to be executed can be triggered by a simple shaking.

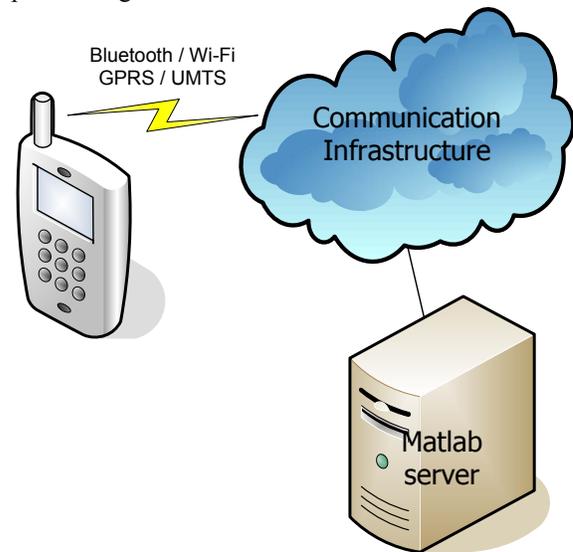


Fig. 1. Overall system architecture

In the following subsections we describe the main components of our system

### B. Image Classification

In this work we use an automatic approach for image indexing and management [2]. Each image in the collection is represented with features related to the presence of faces in the image and features characterizing background and time information. The proposed process of image representation is shown in Fig. 2.

Faces are detected and mapped in an auto emerging space employing eigenfaces. Information from background is managed in a vector space representing low-level features. Time information is normalized according to a reference date. The representation of images is composed by concatenating the three vectors to obtain a single vector.

For each face in the personal album the global representation is given by:

$$x = [x^f, x^b, x^t] \quad (1)$$

where  $x^f$  is the representation of face in the eigenspace of detected faces,  $x^b$  is the background representation for the corresponding image, and  $x^t$  is the time of capture.

A data oriented clustering allows to generate aggregation structures driven by statistical regularities in the represented data.

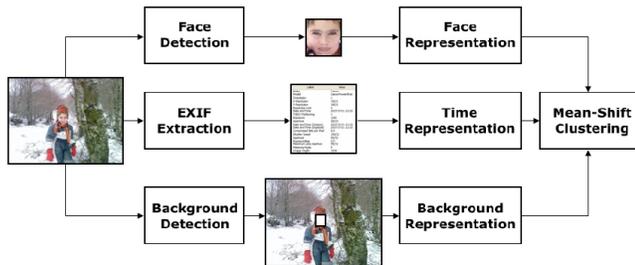


Fig. 2. Image representation for personal photo collections.

1) *Image Representation*: Each image to be archived in the system is searched for faces using the state of the art approach to face detection, that is the framework proposed by Viola and Jones [15]. Several appearance-based approaches could be used for face representation, however eigenfaces [16], i.e., a principal component analysis (PCA) technique, is one of the most mature and investigated method and it performs well while normalizing faces with respect to scale, translation and rotation. Given the set of all detected faces in the collection, each face is described using its representation in the eigenspace.

The largest part of semantic information in personal photo is conveyed in areas where faces appear, the remaining part of the image information is related to the context of the scene. Background information is represented with a composition of color and texture features. In the chosen representation features are globally evaluated and a single vector for each image is produced. Color information is captured through histograms in the RGB color space. Texture is evaluated through Gabor filters [17] with 6 different filters, taking into account 3 orientation and 2 scales.

Temporal data are available for free in image collection through the extraction of EXIF (Ex-changeable image file format) data. The value stored in the time field (i.e., date and hour of capture) is converted in an integer number counting seconds from a fixed data (i.e., Jan 1, 1970). Images are placed in the time line and organized according to time similarity.

2) *Image Clustering*: Faces, background and time information of each image in the collection is automatically organized using a mean-shift clustering technique [18]. Once the input data is organized into clusters we obtain “new collections” of similar images that could be easily browsed and searched.

The clustering process is driven by a set of parameters and although the number of clusters is not fixed, the best bandwidth must be selected. In this work we used an entropy based clustering measure. We choose to evaluate clustering as function of scattering of hand assigned identifiers in the clusters.

Two indexes have been defined. The Intra-Cluster Entropy gives a measure of the entropy inside clusters (i.e., if many different labels are present in a cluster the value of Intra-Cluster Entropy is high, if a label is concentrated in few clusters the entropy has a low value). The second index, the Intra-Label Entropy provides a measure of the distribution of a label across clusters (i.e., if a label is always present, or absent, in a cluster the entropy has a low value, if a label is generally present in many clusters the entropy is high).

To reduce the Intra-Cluster Entropy a lower bandwidth should be preferred, while to reduce Intra-Label Entropy a higher bandwidth should be chosen. To modulate this tradeoff, a Global Clustering Entropy measure has been defined:

$$E_G = \zeta \cdot E_c + (1 - \zeta) \cdot E_l \quad (2)$$

The value of the parameter  $\zeta$  allows to modulate weight of Intra-Cluster Entropy and Intra-Label Entropy in the final clustering.

Information is described as a composition of face representation, time of capture and background representation. To cluster this composite information a multivariate kernel is applied with mean shift procedure. Driven by clustering results, the bandwidth value is automatically chosen and the Global Clustering Entropy is used as performance measure [2]. The process is run for the three domains, and ideally can be applied to all the set of orthogonal feature representing input samples. The merging of clusters among multiple domains is performed as in Comanicu et al. [18].

### C. Mobile Interface to Image Classification

The next step was to enable a Smartphone to act as a remote controller for our image classification algorithm available on a Matlab server, and to show the results.

To this end, we set up a Java MIDlet to browse a list of scripts available as functions to be remotely executed on a Matlab server, to feed the algorithm with a locally or remotely available image, and to show the results (Fig. 3).

The MIDlet thus allows its users to do operations on images, provided that there is a corresponding script on the Matlab server side. The steps leading to the processed image are:

1. send the image to be processed to the Matlab server;
2. remotely invoke the classification algorithm on the image;
3. download the processed image and store it appropriately according to its classification.

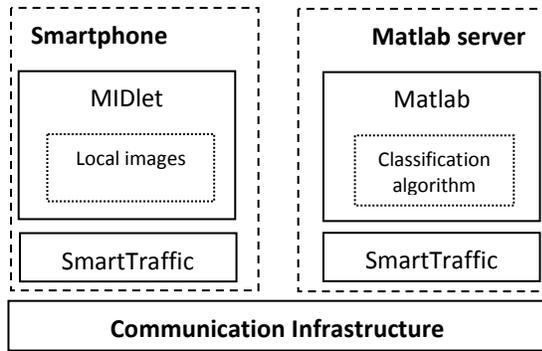


Fig. 3: Layered system representation

This way, should the algorithm be updated, it does not have to be rewritten for execution on mobile devices, since it is executed on the server side as a Matlab script. Furthermore, even complex algorithms can be made available to users, since they do not require any computing power from mobile devices. Actually mobile devices act as remote controllers, and they can be used for other tasks waiting for the server to end the requested image processing.

Once started, the MIDlet automatically connect the Matlab server over the available network connection. The user then choose the local images to be classified and send them to the remote server. The remote server runs the algorithm on the images and sends back the classification terms as in (1). The MIDlet then move the pictures to the corresponding folders and add them some suitable metadata, if possible (depending on the image file type).

#### D. The SmartTraffic Layer

The MIDlet we implemented to enable a smartphone to act as a remote controller is based on a Java-based communication wrapper, called SmartTraffic [4], which allows the MIDlet to seamlessly use both TCP and UDP protocols over both IP-to-IP and Bluetooth-to-IP. This way users can use any type of wireless connection accessible by their Smartphones and available in their neighborhood.

SmartTraffic allows programmers to develop applications providing them with a suitable tool to connect a mobile device with a fixed device, such as a PC, over a wireless connection. SmartTraffic also allows programmers to share the PC internet connection from mobile devices, thus transforming the PC in a proxy server. This feature can be useful to access a Bluetooth-based pervasive framework, where a mobile device “enters” the system through a PC near to it via the Bluetooth connection.

The wrapper provide MIDlets programmers with three network access modes:

- *IP-to-IP*;
- *Bluetooth-to-Bluetooth*;
- *Bluetooth-to-IP*.

People who want to use the SmartTraffic wrapper within their applications, first have to import the appropriate package according to where applications are planned to be run.

## IV. CASE STUDY

In order to carry out our trials, we used three Symbian OS Series 60 smartphones: a Nokia N80 (Series 60v3), a Nokia 5800 XpressMusic and a Nokia N97 (Series 60v5).

We exploited the MatMager MIDlet we implemented in [3], updating it with new interaction modes. Figures 4 a) and b) show two screenshots of our MIDlet running on the Nokia N97. Users must first choose what kind of network access mode they want to use, according to the current availability and to the account charging policy (fig. 4a). Then they can choose whether to take a picture with the embedded camera or pick up an existing one from the phone memory (fig. 4b).

We tried to make the user interface as more straightforward as possible, thus giving the user the idea that the image classification is carried out locally. To achieve this goal, we actually transfer each image only one time, from the mobile to the remote server. The Matlab server sends back only the classification result that is then used by the MIDlet for our purposes.

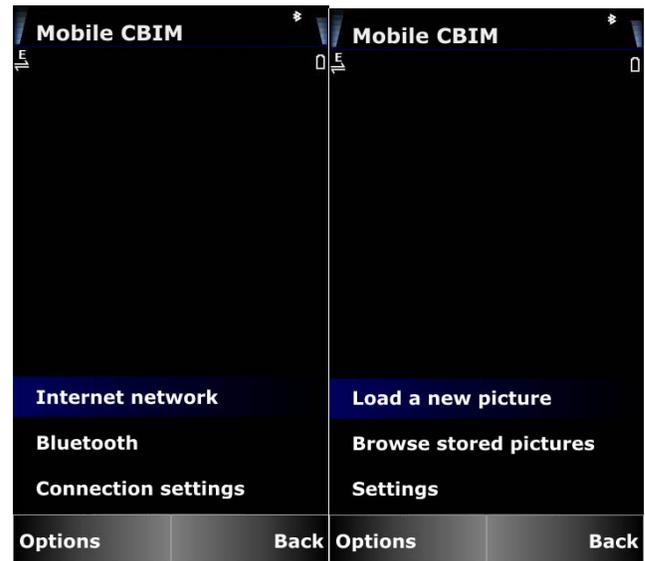


Fig. 4: a) Network access mode selection b) Picture source selection

The image processing layer has been evaluated on a real photo collection of about 1000 images captured by a mobile phone (i.e., VGA and double VGA images) in about three months.

Each image has been manually labeled to store information on the presence of faces, background characteristics and time of shooting. We chose four known people (i.e., each face is defined by an ID), four typical contexts (i.e., indoor, urban, snow, sea) and four events (birthday, trip, holiday, wedding) mainly present in the collection.

The experiments have been aimed to the evaluation of the retrieval capability of the proposed system in terms of faces, background and temporal labeling and an entropy based

analysis of the clustering process has also been performed to better understand the process itself. The results of the classification process are congruent with those showed in [3].

Figures 5 a) and b) show an example of the classification process. In more detail, Fig. 5 a) shows the sample picture we used as key to search for similar ones. Fig. 5 b) shows some of the results retrieved and classified by the algorithm (the first 4 out of 23), and in particular the pictures with the same context as the key one.



Fig. 5: a) Key picture b) Retrieved pictures

In order to give users a more intuitive and funny interaction experience, we added the shake-activation feature to be used on devices that are equipped with accelerometers, and we tried it on our Nokia 5800. In other words, once the key picture is selected, the classification task can be triggered by simply shaking the phone. People who used it, reported this as a very friendly interaction mode, since they make the same moves as when using a sieve.

In order to have an early evaluation of our system, we asked 34 students of our last-year courses to express their opinion by means of an anonymous questionnaire. Table 1 shows the results. The mean satisfaction is high (maybe due to the low age of the sample), except for the response time. We expected that, since users do not know what is the actual data flow from and to their phones. They feel as the classification is carried out locally on their devices, so the response time (from 7.8 to 14.3 seconds during our trials, mainly due to the image transfer over the wireless connection) does not achieve a high rate.

## V. CONCLUSIONS AND FUTURE WORKS

A mobile interface for automatic image searching and browsing on mobile devices has been presented. We considered a photo collection of about 1000 images focusing

on who (faces/people) is in the picture, where (background) and when (time) the picture was shot. A mean-shift clustering technique has been used to manage these three aspects making the indexing process transparent to the user since image features are automatically extracted and clustering parameters are automatically determined according to an entropy based figure of merit.

The MIDlet we used for our trials demonstrated the effectiveness of our design choices, as well as the friendliness of the user interface, with particular reference to the shake-activated one.

Future works will investigate the possibility to run all or part of the classification algorithm on the mobile phone side. Our initial experiments to this end make use of Python-coded applications. This choice allows us to better exploit the mobile phone capabilities, even with the drawback of portability loss.

Table 1: System evaluation by students involved in our tests

	<i>Not useful</i>	<i>A little useful</i>	<i>Quite useful</i>	<i>Very useful</i>
How do you rate the possibility to access a remote image classification service?	0	3	11	20
How do you rate the friendliness of the interface?	0	0	15	19
How do you rate the overall response time of the system? (*)	10	12	8	4
What is your overall rate for the system?	1	6	18	9
(*) trials carried out on a IP-to-IP link over a GPRS-EDGE connection				

## REFERENCES

- [1] Ardizzone, E., La Cascia, M., and Vella, F., "A novel approach to personal photo album representation and management," in [Proceedings of Multimedia Content Access: Algorithms and systems II. IS&T SPIE Symposium on Electronic Imaging], 6820 (2008).
- [2] Ardizzone, E., La Cascia, M., Morana, M., and Vella, F., "Mean shift clustering for personal photo album organization," in IEEE International Conference on Image Processing (ICIP), pages 85–88. IEEE, 2008.
- [3] Ardizzone, E., La Cascia, M., Morana, M., and Vella, F., "Three-domain image representation for personal photo album management," in [Proceedings of Multimedia Content Access: Algorithms and systems IV. IS&T SPIE Symposium on Electronic Imaging], (2010).
- [4] Genco A., Sorce S., Ferrarotto C., Gallea R., Gentile A., Impastato S., Morana M., "A Java-based Wrapper for Wireless Communications", First International Workshop on Intelligent Interfaces for Human-Computer Interaction (IIHCI-2008), Barcelona, Spain, mar. 4-7 2008, pp. 769-774, doi:10.1109/CISIS.2008.129
- [5] L. Zhang, L. Chen, M. Li, and H. Zhang. Automated annotation of human faces in family albums. In Proceedings of ACM International Conference on Multimedia, 2003.
- [6] B. N. Lee, W.-Y. Chen, and E. Y. Chang. A scalable service for photo annotation, sharing and search. In Proceedings of ACM International Conference on Multimedia, 2006.
- [7] J. Cui, F. Wen, R. Xiao, Y. Tian, and X. Tang. Easyalbum: An interactive photo annotation system based on face clustering and re-ranking. In Proceedings of ACM Special Interest Group on Computer-Human Interaction, 2007.
- [8] Apple Inc., Iphone'09. Available at: <http://www.apple.com/ilife/iphone/>.

- [9] Teik-Toe Teoh, Yok-Yen Nguwi and Siu-Yeung Cho, "Intelligent Face Locator for Smartphone", IEEE International Symposium on Industrial Electronics (ISIE 2009), Seoul Olympic Parktel, Seoul, Korea July 5-8, 2009, pp. 1662-1667
- [10] Deok-Hwan Kim, Chan Young Kim, and Yoon Ho Cho. Automatic generation of the initial query set for cbr on the mobile web. In PCM (1), pages 957–968, 2005.
- [11] Karthik Kumar, Yamini Nimmagadda, Yu-Ju Hong, and Yung-Hsiang Lu. Energy conservation by adaptive feature loading for mobile content-based image retrieval. In ISLPED '08: Proceeding of the thirteenth international symposium on Low power electronics and design, pages 153–158, New York, NY, USA, 2008. ACM.
- [12] I. Ahmad, S. Abdullah, S. Kiranyaz, and M. Gabbouj. Content-based image retrieval on mobile devices. In Proceedings of SPIE (Multimedia on Mobile Devices), 2005.
- [13] J. S. Hare and P.H. Lewis. Content-based image retrieval using a mobile device as a novel interface. In Storage and Retrieval Methods and Applications for Multimedia, 2005.
- [14] M. Gabbouj, I. Ahmad, Malik Y. Amin, and S. Kiranyaz. Content-based image retrieval for connected mobile devices. In Proceedings of Second International Symposium on Communications, Control and Signal Processing (ISCCSP), 2006.
- [15] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2001.
- [16] M. Turk and A. Pentland. Eigenfaces for recognition. Journal of Cognitive Neuroscience, 3(1):71–86, 1991.
- [17] A.K. Jain and F. Farrokhnia. Unsupervised texture segmentation using gabor filters. In Systems, Man and Cybernetics, 1990. Conference Proceedings., IEEE International Conference on, 1990.
- [18] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. IEEE Transaction on Pattern Analysis and Machine Intelligence, 24:603–619, May 2002.
- [19] Weiser M.: The Computer for the Twenty-First Century, Scientific American, pp. 94-100, (1991).
- [20] Berhe G., Brunie L., Pierson J.M., "Modeling Service-Based Multimedia Content Adaptation", Proceedings of the 1st ACM Conference on Computing Frontiers CF '04, Ischia, Italy 2004, pp. 60-69
- [21] Genco A., Sorce S., Reina G., Santoro G., "An Agent-Based Service Network for Personal Mobile Devices", IEEE Pervasive Computing, vol. 5, no. 2, Apr-Jun 2006, pp. 54-61
- [22] Pilato, G., Augello, A., Santangelo, A., Gentile, A. and Gaglio, S.: An Intelligent Multimodal Site-guide for the "Parco Archeologico della Valle dei Templi in Agrigento, Proc. of First European Workshop on Intelligent Technologies for Cultural Heritage Exploitation at The 17th European Conference on Artificial Intelligence, Riva del Garda, Italy, August, 45-49 (2006)
- [23] Price, S.; Summers, R., "Mobile Healthcare in the Home Environment", Proceedings of 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBS '06, New York, NY, Aug. 2006, pp. 6446-6448
- [24] Raptis D., Tselios N., Avouris N., "Context-based design of mobile applications for museums: a survey of existing practices", Proceedings of the 7th ACM International Conference on Human-Computer Interaction with Mobile Devices & Services, Salzburg, Austria 2005, pp: 153-160
- [25] Henrysson A., M. Ollila, M. Billinghurst, "Mobile phone based AR scene assembly", Proceedings of the 4th International Conference on Mobile and Ubiquitous Multimedia, Christchurch, New Zealand 2005, pp. 95-102
- [26] N.K. Vuong, S. Chan and C.T. Lau, "Classification of pH Levels using a Mobile Phone", Proceedings of the 13th IEEE International Symposium on Consumer Electronics (ISCE2009), pp. 823-827
- [27] Genco A., S. Sorce, G. Reina, G. Santoro, R. Messineo, R. Raccuglia, L. Lo Vecchio, G. Di Stefano, "An Augmented Campus Design for Context-aware Service Provision", Proceedings of the 33rd annual ACM SIGUCCS Conference on User Services, Monterey, CA, Nov. 6-11 2005, pp. 92-97