A Portable Content Based Image Retrieval System for Art Galleries and Museums.

Giuseppe Conti, Paolo Bincoletto, Giuliana Ucelli

Graphitech, Salita dei Molini,2, 38050 Villazzano (TN), Italy
{giuseppe.conti,paolo.bincoletto,giuliana.ucelli@graphitech.it}

Abstract

In the past decades personalised content delivery has been only achieved through traditional audio-guides. These feature a very primitive form of interaction, mostly based on a keypad where the user selects a number corresponding to the exhibit he/she wants to be informed about. It is clear that such limitation can be clearly overcome by the adoption of modern multimedia-based applications. Furthermore recent development of mobile hardware, such as the last generation mobile phones or Personal Digital Assistants (PDA), it gives, for the first time, the chance to exploit notable processing power available on very light and portable devices, that few years back was only available on standard personal computers.
The work presented in this paper fosters such novel approach by proposing an interactive portable system that supports the user during a visit to a museum by providing customised multimedia data.

Introduction

The fruition of museums, exposition or art galleries can be greatly enhanced when the relevant information is provided to the visitor. More often than not in fact the correct interpretation of an exposition's content requires a deeper knowledge of the exhibits' context. Often additional information is necessary to frame the exposition within the proper historical, cultural or technical perspective. Furthermore this information, in order to best tailor each visitor's personal background, should be delivered on a personalised base, ideally involving the visitors themselves in an interactive and exciting process.
The work presented in this paper shows the system developed by the authors to tackle this issue. The result of the research has brought to a complete application capable of supporting the visitor during the visit through the playback of multimedia information regarding the exhibit.

In the first part of the paper we introduce the issue of the need for mobile IT-based tools in museums and exposition. Then a description of existing state of the art shows current trends and applications in the field of mobile support to exhibitions' visitors. Then the system developed is introduced and both the authoring platform and the user tool are illustrated. The paper will illustrate the details of the system, it will present the algorithm and the modifications which have been introduced to deal with the limited resources of the hardware available and it explains its main strengths and weaknesses. Further the paper presents a test run to assess the robustness of the system under several working conditions. The results of the test will be discussed. Finally the paper explains the system's limitations and illustrates the future improvements to the current system.

The Adoption of IT-based Technologies in Museums and Exhibitions.

As discussed in previous works [1] the chance of a first person experience of an exhibition can have great value on people's learning process. For this reason the availability of an efficient customizable guiding represents a key feature for the best possible delivery of information through the exposition's content. Nevertheless most guiding systems are clearly limited. The majority of these in fact are based on standard audio guides, intrinsically limited in the nature and extent of the content which these are capable to convey. To overcome this limitation nowadays computers, touch-screens, video projectors are more and more adopted in order to overcome the limit of the audio guides and to deliver to the final user media-rich information right on the exhibits. Further this approach allows for a higher level of interactivity which in turn fosters a more engaging experience and hence more effective learning process.

However the main limitation of this approach can be found in the limited level of customization that can achieved since the content is defined for a single type of user. Furthermore a single user at time can benefit from such interactive experience whilst the other participants are just relegated to the mere role of observers.

The approach proposed in this research work tries to tackle the aforementioned issues by promoting use of portable IT technologies in museums and expositions. The work presented proves how a PDA-based (Personal Digital Assistant) system, can be fruitfully used to enhance the exposition's content and how this approach can enhance the learning process by providing clear, comprehensive, easy-to-use personalised multimedia content.

Related works

As described in the literature [2] [3] [4], the birth of traditional audio guide dates back to some 40 years ago. However its first 35 years of history have seen only a slow and limited evolution until the major shift from analogical to digital took place. These devices now support comments in various languages of the duration of beyond a hundred of hours. The interaction is limited to audio comments in response to keyboard's commands. Some devices [5] support the possibility to activate the comments through detection of radio waves or infrared beams. Other commercial systems are based on PDAs [6] [7] [8] where some multimedia content is played back after the insertion of a code corresponding to the exhibit or according to the positioning of the device.

Research works, on the other hands, are currently trying to push this approach further by proposing complex platforms capable of advanced functionalities on standard PDA devices. Example of this tendency is the PEACH (Personal Experience with Active Cultural Heritage) project [9] whose aim was to create a personalized and interactive guide, with the aid of advanced technologies which can improve the appreciation of the cultural assets. Infrared beamers mounted on the ceiling of the exhibition's premises identify the user position and a video clip is started every time the user stands in front of an exhibit.

The ARCHEOGUIDE (Augmented Reality-based Cultural HEritage On-site GUIDE) project [10] aimed at the delivery of a system that could provide new form of access to information on archaeological sites. The intention of the project was to stress particularly the virtual reconstruction of the ruins, while providing information on the archaeological site being visited. The ARCHEOGUIDE backbone is made of a central server, which stores the database with all information on the site. However the system is required to carry a portable unit made of a lightweight portable computer, a HMD (Head Mounted Display), a camera, a microphone and headsets. The portable unit is then connected to the main server through a wireless LAN. The AMIRE (Authoring MIxed Reality) project [11] allowed non-expert users to use Mixed Reality for their applications. A case study carried on at the Guggenheim museum in Bilbao showed how it was possible to show information through augmented reality techniques on the exhibit once the tracking system recognized appropriate marker close to the exhibit itself.

The PhoneGuide project [12], developed at the Bauhaus University, Weimar (Germany), in collaboration with the Natural History Museum of Frankfurt, aimed at creating one of the first prototypes of guide for tourist in a museum context. The application is implemented on cellular telephone with integrated camera. Sophisticated functions carry out template and pattern-matching to extract information about the object being pointed at. The system has been adapted to the limited resources available with the chosen device.

As far as the image retrieval algorithms are concerned a number of approaches are found in literature. As observed by Deb & Zhang [13] in the last decade research has been undertaken with the aim to find efficient of image retrieval techniques. In particular, focusing our interest on CBIR (Content-Based Image Retrieval) algorithms, these have proved to be extremely useful in the query of multimedia databases. These algorithms are based on the automatic extraction of a number of features from an image based on colour, texture or shape [18]. In particular the development of such methods for portable devices, which are tipically characterised by very restricted processing power, has been subject of a number of studies.

Chen et al. [14] have developed a mobile wireless device called Butterfly-Watching Learning (BWL). The system, designed for outdoor mobile learning, or *m-learning*, uses a PDA fitted with a CCD camera. The teacher uses a standard notebook fitted a Wi-Fi LAN card, which acts as local server containing a database of images of butterflies. Every student takes a photo to a butterfly, and the PDA transmits the image to the server. Here CBIR techniques are adopted in order to search the information tied to the corresponding butterfly. The information is then sent in real time back to the PDA client. Yu et al. [15] propose an analogous system for outdoor learning which instead recognises beatles. Other systems propose similar approaches for different contents [16] [17].

Description of the System

The system developed is extremely simple and intuitive to use, since it only requires a simple and natural gesture of the visitor who has to point the PDA to the artefact he/she wants to be informed about (see Figure 1).

Figure 1: example of usage of the system

The system is very lightweight, truly portable, it does not require any additional sensor (infrared, radio, ultrasound, optical etc.) to be installed at the exhibition premises making it quite suitable also within cultural heritage scenarios. In fact The system is based on a standard PDA handheld device fitted with a video camera and with an optional wireless headset. As soon as the user points the PDA's camera to an exhibit this is visualized on the display and the picture is automatically processed by an image processing system which extracts a set of relevant features. A matching algorithm, optimised to take into account the limited hardware resources available through the PDA's CPU, compares the image with those previously stored in a database within the PDA memory. When the system finds the database entry it plays the relevant multimedia content such as audio, text etc. (see Figure 2).

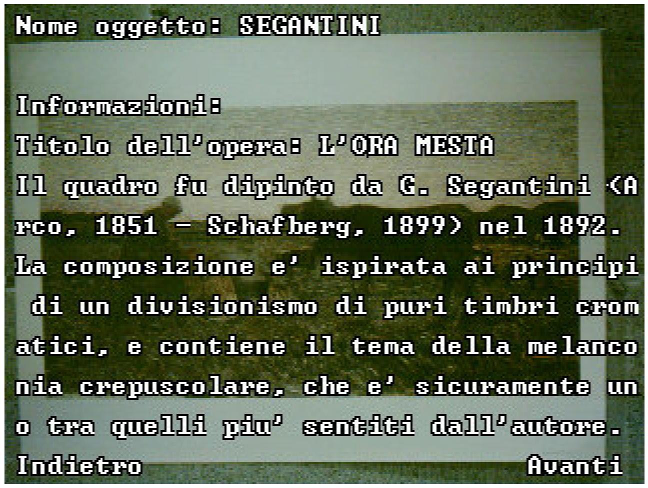

Figure 2: example of playback of multimedia data after the image has been recognised from the database.

Staff from the exhibition can use the system to author the multimedia content and to easily create the database required for the retrieval process (see Figure 3).

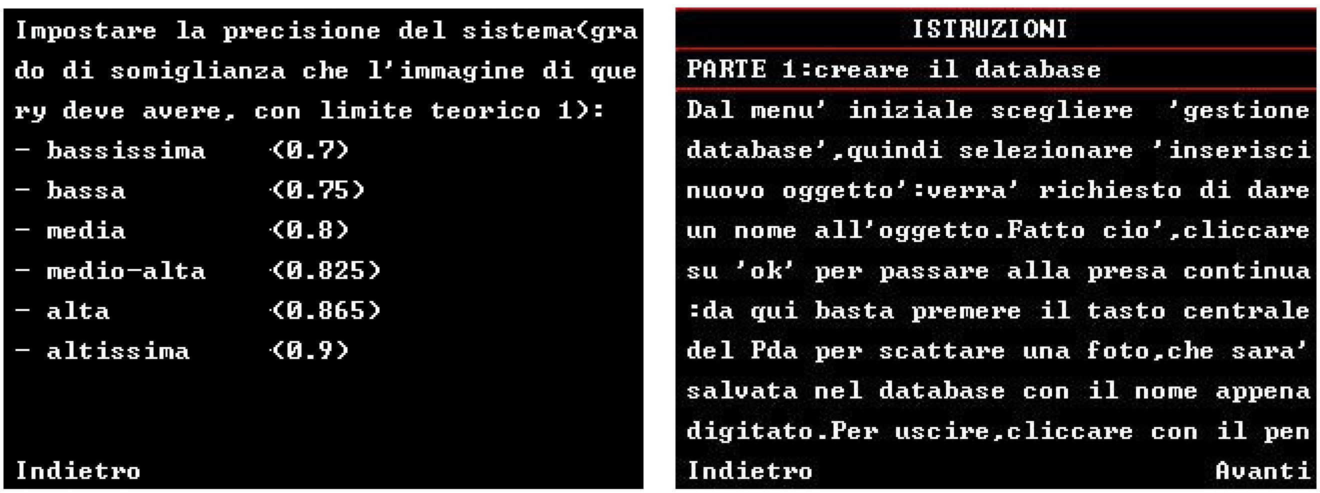

Figure 3: The interface allows to control the main functions during the authoring process

Details of the Algorithm Adopted

The method adopted for the image retrieving process belongs to the so-called CIBR (Content-Based Image Retrieval). These methods differ from traditional methods of image indexing in the fact that they do not require indexing of each image present in the database through keywords or IDs, usually manually assigned. Authors [18] place a great emphasis on the advantages of CBIR methods over traditional ones both in terms of reliabilities and precision.

With regards to the choice of the CBIR method to be adopted, we have taken into account the trade-off between, on the one hand, reliability and precision and, on the other, computational load. For this reason we have not take into account methods relying on shape, segmentation or texture, given their greater demand, in terms of calculation time required by such algorithms. We excluded also algorithms based only on colour analysis, as proposed from Swain & Ballard [19], since they not yield sufficient precision and accuracy.

From these considerations, it has been decided to adopt a method CBIR belonging to the category based on the analysis of colour and space and in particular to pursue the approach suggested by Cinque et al. [20]. The main advantages deriving from such decision are the following:

1. The class of algorithm based on colour and space on one side offers better trade-off between reliability/precision and computational requirements.
2. The algorithm performs satisfactorily, in line with other algorithms such those proposed by other authors [21] [22] [23].
3. The method does not involve complex calculations, or small overhead if compared with [21][22].
4. The approach does not require setting up numerous parameters as required in [21].

The technique described in [20], upon which we based our development, is called SCH (Spatial Chromatic Histogram) method, and it consists in the extraction of features containing information on the pixels' colour and on their location and arrangement within the image. For every image contained in the database we have then calculated its SCH offline and storered these parameters in the PDA memory. When the user acquires a picture through the camera of the PDA, this is used as the query image. The algorithm then extracts its features and calculates the new SCH online. Further the system compares the resulting value with the SCH previously stored in the memory, and returns the highest score of similitude obtained via this comparison. In the following pages we will describe in detail how the SCH of a generic image is calculated, and then how the values of similitude between two images are generated.

Given an image *I*, $N \times M$, with *N* columns and *M* rows, we first reduce its number of colours re-quantizing the image by a fixed palette of *c* colours; in our system we used five palettes:

$$c \in \{9,27,36,64,166\}.$$

Then we may define the colour set:

$$A_k^I := \{(x, y) \in I : I[x, y] = k\}$$

and:

$$h_I(k) = \frac{\left(\left|A_k^I\right| * 100\right)}{N \times M}$$

as the normalized histogram of the c colours in the image I. This is the first of the three features that form the SCH parameter. While h stands for the colour information of the image, the other two features represent, in an approximated way, its spatial information, that is the location and the spatial arrangement of the pixels of the same colour within I.
The second feature is the centre of mass defined by:

$$\bar{b}_I(k) := (\bar{x}_k, \bar{y}_k);$$

where:

$$\bar{x}_k = \frac{100}{N}\frac{1}{|A_k|}\sum_{(x,y)\in A_k} x;$$
$$\bar{y}_k = \frac{100}{M}\frac{1}{|A_k|}\sum_{(x,y)\in A_k} y.$$

Finally the last feature is the standard deviation:

$$\sigma_I(k) = \sqrt{\frac{1}{|A_k|}\sum_{p\in A_k} d\left(p, \bar{b}_I(k)\right)^2},$$

where p is a generic pixel of coordinates:

$$x_{p_1} = \frac{(x*100)}{N},\ y_{p_1} = \frac{(y*100)}{M}.$$

in A_k^I and $d\left(p, \bar{b}_I(k)\right)$ is the Euclidean distance between that pixel and the corresponding centre of mass with the same k fixed.
Finally, the SCH $\bar{S}_I(k)$ of the image I is defined as a vector of these three features:

$$\bar{S}_I(k) = \left(h_I(k), \bar{b}_I(k), \sigma_I(k)\right),$$

with $1 \le k \le c$ to identify each colour. Having calculated the SCH for each image in the database, given a query image Q and a generic database image I, the system calculates also:

$$\bar{S}_Q = \left(h_Q, \bar{b}_Q, \sigma_Q\right)$$

for the query image, and it compares the SCH of Q and I through the following similitude function:

$$f_s(Q,I) = \sum_{i=1}^{c} \min\left(h_Q(i), h_I(i)\right) * \left(\frac{100*\sqrt{2} - d\left(\bar{b}_Q(i), \bar{b}_I(i)\right)}{100*\sqrt{2}} + \frac{\min\left(\sigma_Q(i), \sigma_I(i)\right)}{\max\left(\sigma_Q(i), \sigma_I(i)\right)}\right)$$

However, unlike proposed in [20], where the result of the retrieval consists in a ten of images extracted from the database, ordered according to the level of likeness with the query (that is according to the value of the function) we have decided to extract the single most similar image. As we'll show in the results, we have also fixed a threshold *S* setting the minimum value of similitude to be accepted. This means that if:

- $f_S > S$, then the systems outputs the name and multimedia content related to the recognized image;
- $f_S \leq S$, the system doesn't recognize the object; therefore the user is invited to take another (possibly better) picture of the exhibit.

The decision to use a threshold was made to avoid occasional errors arsing from not well-focused pictures, or pictures of a non-database image, etc.

In order to verify the robustness of the system, and to be able therefore to estimate the opportunity to carry out its porting on a PDA, first implementation of the algorithm was developed on a standard PC. This was tested creating a database of several images hung to the walls of a room to simulate the pictures of a museum.

As mentioned before the key factor in the development the application has been the limitation imposed by the hardware sets up which has influenced not only the calculation of the parameters of the SCH, but also their representation. In fact the normalisation of the features required the use of real numbers. However, the management of floating point numbers through the PDA processor had revealed rather inefficient. In fact, with a first code in C++ that employed the types `double` and `float`, the processing time required to calculate the SCH for a single image proved to be extremely long, in the order of 40+ seconds for an image with only c=27 (colours). For such reason, it had become necessary to provide an alternative representation which made use of entire numbers of the type `int`. From what said, it was clearly unfeasible proposing a simple conversion of type (casting) of the C++ code of the program. As a result, we abandoned the use of the types `double` and `float`, to which the CPU it dedicates respectively 8 and 4 byte. The equations previously detailed have been modified to take into account this factor.

It is worth noting that dealing with normalisation requires handling possible *over-flow* issues (i.e. not to exceed the maximum manageable number with a sure type of data). In this case the a priori knowledge of the dimensions of the images has ensured possible *over-flows*. Since in fact we have kept fixed N = 320, M = 240, in the improbable event of a monochromatic image, a total 76800 pixel would have the same colour, therefore this would yield to:

$$\left|A_k^I\right| * 100 = 7680000 \cong 7 * 10^6$$

$$\sum_{(x,y) \in A_k} x = \sum_{i=1}^{M} \sum_{j=1}^{N} x = M \sum_{j=1}^{N} x = \frac{M(N+1)N}{2} = 12326400 \cong 10^7$$

$$\sum_{(x,y)\in A_k} y = \frac{N(M+1)M}{2} = 9254400 \cong 9*10^6;$$

Finally it can be proved that:

$$\sum_{p\in A_k} d\left(p_1, \bar{b}_I(k)\right)^2 = 12797552 \cong 10^8..$$

It is therefore clear that the adoption of a type capable of a maximum value higher than the previous figure would have been enough for our purposes. For this reason, since 4 Bytes can represent approximately up to $4*10^9$, we chose to adopt a `unsigned int`.The choice of the type `unsigned int` has brought a remarkable contraction of the necessary calculation times, as described in the following section.

The Key Study and its Results

In our test we simulated the room of a museum, by placing some paper paintings on the wall of our laboratory. We have run the experiment under controlled artificial light conditions. Each picture was shot frontally, it contained the entire object, and it followed the heuristic rule of 'background dimensions << object dimensions'.

By testing the system with 10, 15 and 20 paintings, setting the threshold $S = 0.7 f_s$ and using all the five available palettes, we obtained the results shown in the graphics below, where the term precision stands for the percentage of correct recognition of an image. The results were most encouraging, with the accuracy of the system always at a very good level as showed below where the results are plotted.

As foreseen the best result were achieved with simpler palettes which generated little or no error during the retrieval process. Furthermore this has the advantage to yield a much faster system capable of reacting quickly to the input of the users.

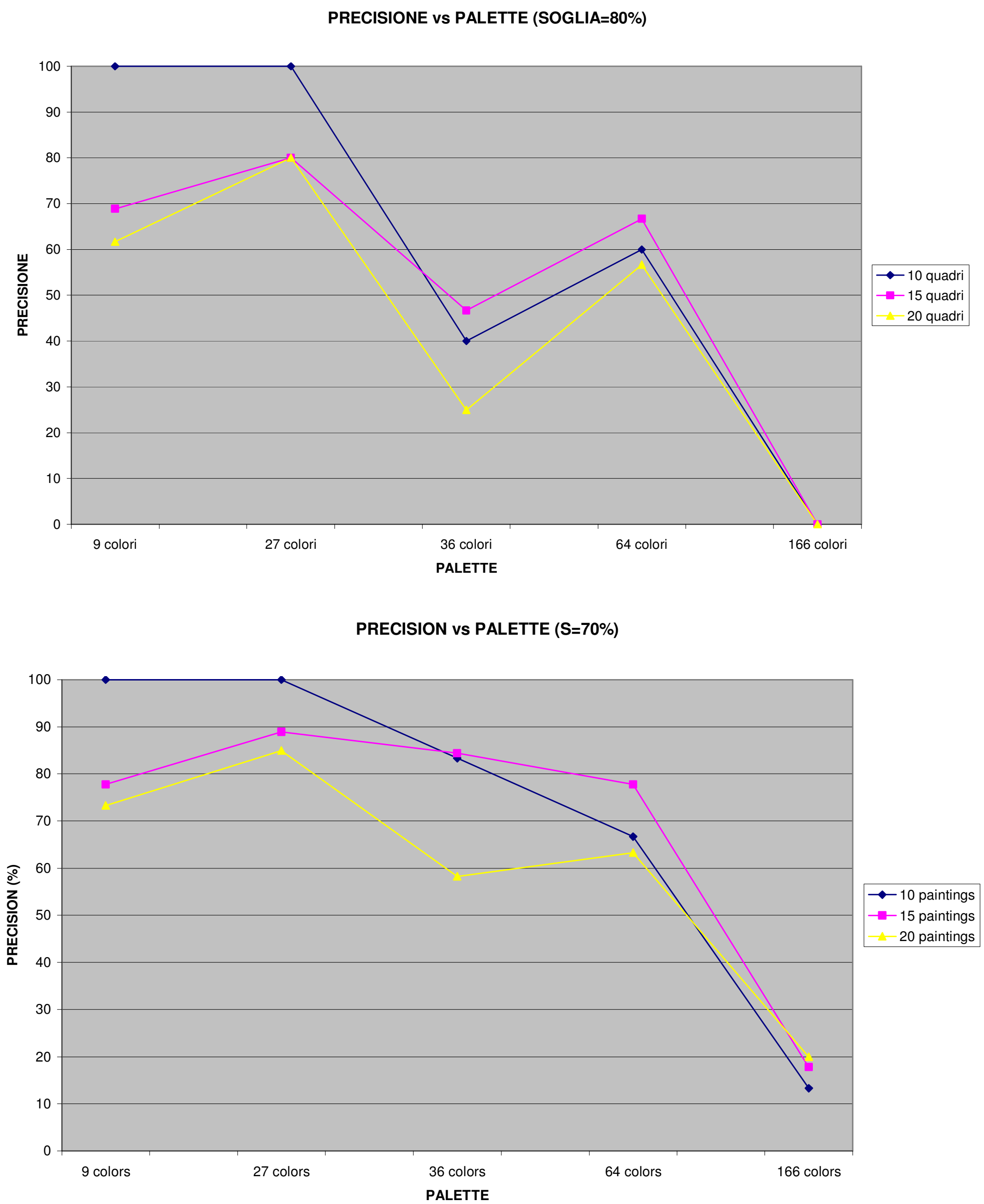

Figure 4: Precision of the system in function of the palette, for 10, 15 and 20 paintings, with S=70% (below) and S=70% (above)

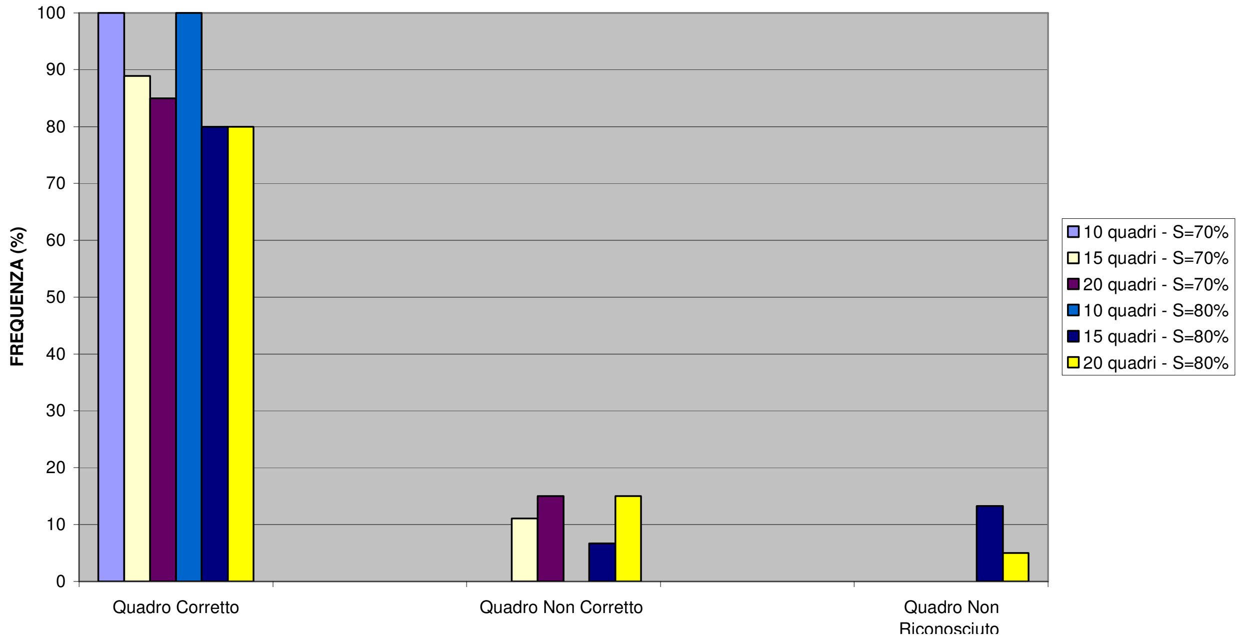

Figure 5: frequency of results for the 27 colors palette (the most performing) when both the threshold and the number of paintings varies. On the left are grouped the correct hits, in the middle the wrong recognitions, on the right missed recognitions.

The table below reports the frequency of correct, non correct recognition (if $f_s > S$) or non recognition (if $f_s \leq S$), with varying threshold and number of paintings. It shows the results only for c=27, which in our test proved to be the best choice of palette:

RECOGNITION FREQUENCY (%) – PALETTE OF 27 COLORS			
Number of paintings and threshold	Correct painting	Non-correct painting	Non-recognized painting
10 paint. - S=70%	100	0	0
15 paint. - S=70%	88.9	11,1	0
20 paint. - S=70%	85	15	0
10 paint. - S=80%	100	0	0
15 paint. - S=80%	80	6,7	13,3
20 paint. - S=80%	80	15	5

Figure 6: Frequency of the recognition results for the palette of 27 colors, with varying threshold and number of paintings

Conclusions and Further Works

The performances of the system have been generally quite good. The advantages of our system are, conforming to the requirements in Träskbäck et al. [24], those of using non-intrusive technology, of being light, easy to carry, and sociably acceptable. It also has got a simple and intuitive user-interface, and it does support multimedia audio-textual messages. The cost of the whole system, being that of the PDA plus a micro-camera, is also quite cheap, resulting in a good trade-off between provided services and cost.
The main advantages are therefore:

- Good accuracy
- Use of not-invasive technology
- Reduced costs
- Ease of transportation, lightweight easily acceptable by users
- Automatic functioning
- User friendliness (intuitive UI, easily upgraded and re-used, expandable with further features)
- Multimedia support

The bad side of the system, anyway, are that of its robustness related to the retrieval algorithm (in fact, the system may fail the recognition), and that of its speed, due to the limited PDA hardware, which also doesn't allow to implement too complex CBIR techniques. On overall we can say that the main limitations are:

- Optimal functioning only under certain conditions
- Sub-optimal real-time behaviour (≈2 secs for retrieving) due to limited HW performances
- Limited operating time due to PDA's battery capacity

Therefore, to improve the system's performances, we would suggest the testing and implementation of other CBIR algorithms, in parallel with the use of faster and more powerful hardware.

References

[1] Conti, G., Bincoletto, P., De Amicis, R. (2005). A Mobile Information System for Exhibitions and Museums. In Methods and Technologies for Learning, G.Chiazzese, M. Allegra, A.Chifari, S. Ottaviano (Eds.), WIT Press. 2005, pp.355-360 ISBN 1-84564-155-8.

[2] Tellis, C. & Proctor, N. (2002) Workshop: Handhelds in Museums, Museums and the Web, Boston, 2002.

[3] Tellis, C. (2001). Panel Presentation, Museum Computer Network Conference, Cleveland, November 2001.

[4] Proctor, N. & Tellis, C. (2003). The State of the Art in Museum Handhelds in 2003. Tratto dalla conferenza Museums and the Web 2003. Antenna Audio, United Kingdom and USA. Reperito il 2 ottobre 2004 dal World Wide Web: http://www.archimuse.com/mw2003/papers/proctor/proctor.html

[5] Clicsound Audioguide. http://clicsound.com/

[6] Antenna Audio. http://www.antennaaudio.com/

[7] Ophyris Système (2004), audioguide. http://www.ophrys.net/

[8] Flick Software (2004). Mobile Interactive Guide http://www.flicksoftware.com/
[9] PEACH (2004) project. http://peach.itc.it/home.html
[10] ARCHEOGUIDE (2004) project http://archeoguide.intranet.gr/
[11] AMIRE (2004 project http://www.amire.net/
[12] PhoneGuide (2004). Computer Vision Library for Mobile Phones. http://www.uni-weimar.de/~bimber/research.php
[13] Deb, S. & Zhang, Y. (2004). An Overview of Content-based Image Retrieval Techniques. 18th International Conference on Advanced Information Networking and Applications (AINA'04) Volume 1, pp. 59-65.
[14] Chen, Y.-S., Kao, T.-C., Yu, G.-J. & Sheu, J.-P. (2004). A Mobile Butterfly-Watching Learning System for Supporting Independent Learning. 2nd IEEE International Workshop on Wireless and Mobile Technologies in Education (WMTE'04).
[15] Yu, G. J., Chen, Y.-S., Shih, K.-P. (2004). A Content-Based Image Retrieval System for Outdoor Ecology Learning: A Firefly Watching System. 18th International Conference on Advanced Information Networking and Applications (AINA'04) Volume 2, p. 112-115.
[16] Noda, M., Sonobe, H., Takagi, S. & Yoshimoto, F. (2002). Cosmos: Convenient Image Retrieval System of Flowers for Mobile Computing Situations. Proceedings of IASTED International Conference, Information Systems and Database (ISDB), pp. 25-30.
[17] Sonobe, H., Takagi, S., Yoshimoto, P. (2004). Mobile Computing System for Fish Image Retrieval. Informazioni reperite l'1 ottobre 2004 dal World Wide Web: www.wakayama-.ac.jp/~saeko/docs/sonobe_IWAIT2004.pdf
[18] Eakins, J. P. & Graham, M. E. (1999). Content-based Image Retrieval: A report to the JISC Technology Applications Programme http://www.unn.ac.uk/iidr/report.html
[19] Swain, M. J. & Ballard, D. H. (1991). Color Indexing. Da Inernational Journal of Computer Visions, 7(1), pp. 11-32.
[20] Cinque, L., Levialdi, S., Olsen, K. A. & Pellicanò, A. (1999). Color-Based Image Retrieval Using Spatial Chromatic Histograms. IEEE International Conference on Multimedia Computing and Systems Volume II-Volume 2, pp. 969-973.
[21] Chua, T. S., Tan, K.-L. & Ooi, B. C. (1997). Fast signature-based color-spatial image retrieval. 1997 International Conference on Multimedia Computing and Systems (ICMCS '97), pp.362-369.
[22] Liu, Y. & Ozawa, S. (2000). A Integrated Color-Spatial Image Representation and the Similar Image Retrieval. 4th IEEE Southwest Symposium on Image Analysis and Interpretation, pp. 283-287.
[23] Rao, A., Srihari, R. K. & Zhang, Z. (1999). Spatial Color Histograms for Content-Based Image Retrieval. 11th IEEE International Conference on Tools with Artificial Intelligence, pp.183-186.
[24] Träskbäck, M. & Nieminem, M. P. (2003). Requirements for Using Mixed Reality in Museums. First Research Workshop on Augmented Virtual Reality (AVIR 2003).