SUPPORTING CULTURAL EMOTIONAL BROWSING FOR MUSEUMS:
THE VERSOVERDI APP

M. Roccetti, G. Marfia, C. Bertuccioli, A. Marcomini, M. Zanichelli, A. Varni
Department of Computer Science and Engineering
Alma Mater Studiorum – University of Bologna
Mura A. Zamboni 7, 40127 Bologna, Italy

ABSTRACT

Among the infinite options that are only limited by one’s creativity, many different applications are flourishing in the context of cultural heritage. In fact, while cultural heritage is often hard to understand or even only to reach, technology is today providing all with a powerful access key to many of its very different instances. Such key can be based on traditional means of communication and explanation (e.g., audio guides), or also ambitiously exploit novel ways of representation and visualization (e.g., augmented reality renderings). Our contribution in this area is that of designing and implementing a mobile app that supports the navigation of Giuseppe Verdi-related cultural sites. Unlike any other app, ours proposes an innovative human-computer interaction scheme that supports mood-based browsing of cultural information across multiple museums at once. In practice, our app is capable of determining a user’s mood and, based on such information, of providing visit suggestions as well as accessibility to content that matches the given mood state.

Index Terms— Emotional browsing, human-computer interfaces, cultural heritage museums.

1. INTRODUCTION

The opportunities that can rise from the meeting of two different worlds, information technology and cultural heritage, are incredibly vast and mostly unexplored. While technology provides the opportunity of virtually reaching everyone everywhere, utilizing means of communication that are becoming widely understood and recognized, the many different forms of cultural heritage provide an unimaginable amount of content and stories that are rightfully worth of being presented and reinterpreted.

Digital technologies, in particular, have pushed their way to assume the role of being one of the great mediators between humanity and reality. All those places that cannot be seen (e.g., Europe during the Renaissance), that are hard to be reached (e.g. Antarctica), and that could also be impossible to imagine (e.g., the planet Earth in 5000 years), can now be grasped thanks to their use. Interestingly, the role that digital devices are today playing is very similar to the one played for centuries by literature and, more recently, for decades by cinematography. Far away countries could only be imagined through the lines written by explorers and novelists, or seen thanks to the pictures and movies shot by photographers or film directors. However, unlike traditional literature and cinematography, contemporary digital systems can serve many different purposes (e.g., phone, document editor, TV-set, remote control, etc.) and pervasively aid their users in any place, being one of the factors that today influence most how social and human-reality interactions occur. All this is confirmed by our daily experience: digital technologies are slowly, but undauntedly finding their way in all of the domains of interest to human beings.

Public exhibitions and museums represent no exception, as a central role is already being played by technology in the related fields of entertainment and education. However, just as computer gaming has attracted more people, in the past few decades, with the adoption of new and natural interface systems (e.g., Wii, Kinect) and mixed reality technologies than with the increase of CPU processing power, we observe that in the context of a museum, technology can be similarly used to create an invisible and instinctive link between its visitors and its exhibits. The problem, hence, is how interaction designers can effectively create such type of relation.

In such scenario, computer professionals are endowed with a privilege and a responsibility, whose fulfillments constitute the foundations to the successful design of new and advanced Human-Computer Interface (HCI) systems. The privilege is given by the opportunity of shaping how human beings interact with the surrounding world, thus being able to create new and intuitive ways in which fast exchanges of information occur, as well as providing multiple and completely new experiences and perspectives of reality. The responsibility, instead, involves responding to the needs of an audience that is often composed of computer illiterates, but nonetheless eager to benefit from the use of digital systems. On top of these foundations, a fortunate HCI cannot be constructed without a thorough
comprehension of the peculiarities of the scenario where it will be employed.

Now, many different museums have been working at creating new means of access in order to provide their visitors with a better comprehension of the treasures that they preserve. This process, however, is very often carried out in isolation by single museums, as their curators usually pose most of their attention on their own exhibits. Nevertheless, numerous connections and links can often be found among exhibits residing in different institutions. This fact poses an interesting problem: how can we support an efficient and playful browsing of all the content and information that is typically spread out among different institutions?

A great opportunity of working on such problem has been provided by the Istituto per i Beni Artistici Culturali (IBC) dell’Emilia Romagna. This institution, which manages more than five hundred museums in the Emilia Romagna region, is celebrating the birth of Giuseppe Verdi, the famous opera composer. Such celebration is considered particularly important by IBC, as Verdi was born in Emilia Romagna and, due to this fact, IBC runs many museums and sites that are devoted to the preservation and promotion of his memory and work. Hence, in such scenario, we have been requested to create an amusing way of browsing the content that many different museums preserve of Giuseppe Verdi’s life, with the twofold objective of encouraging people to: (a) discover the links that connect those museums and, ultimately, (b) reach and visit them.

To answer such request, we designed and implemented a new mobile app, VersoVerdi (its icon is shown in Fig. 1), whose original contributions are those of:

1. Supporting emotional browsing: VersoVerdi recognizes the mood (e.g., happy/unhappy) of a user and, based on such information, provides content that well matches that mood;
2. Suggesting links between museums. To this aim, we graphically represented every museum as a universe of planets where, following such metaphor, each planet suggests the routes that lead to other related ones.

In essence, unlike any other mobile app, VersoVerdi implements a binary emotion recognition mechanism which is capable of recognizing whether its user is smiling (and hence interpreted as happy) or not (i.e., unhappy) implementing algorithms that solely rely on local resources, without resorting to the processing power offered by any remote server. Once such piece of information is unveiled, VersoVerdi unlocks videos, audio tracks and pictures, which match the given mood state, subsequently suggesting new paths of exploration of cultural heritage content and sites.

This paper is organized as follows. In Sections 2 we provide an overview of the apps that fall closest to the spark that animates VersoVerdi. We then concentrate on an overview of the architecture of our app in Section 3 and on its evaluation in Section 4. We finally conclude with Section 5.

![Fig. 1. The VersoVerdi app universe.](image)

## 2. RELATED WORK

Although a wealth of research has so far been carried out in the field of mobile apps, a fairly limited number of applications have been so far designed posing their emphasis on the cultural world. Among those that fall into this domain, we will here focus on those that are closest to our approach.

Artist’s View and Marsili’s Spirit, two mobile apps that recently appeared on the Apple App Store, have one element in common: they both support semantic and geographic connections between different locations and museums [1], [2]. In particular, the peculiarity of Artist’s View is that of providing its users with the locations, and hence with the views that inspired the painters that portrayed Florence at the end of the nineteenth century. While touring through Florence, a user of Artist’s View is given the chance of being alerted when one of the given viewpoints is reached and, when this happens, of juxtaposing a copy of the related painting to the scene that today appears in front of him/her. Clearly, Artist’s View is limited by the fact that it focuses on the exhibits of a single institution (e.g., Palazzo Strozzi in Florence), without unveiling the relations that those paintings, and its authors, probably had with many other places in Florence.

Marsili’s Spirit, instead, is an app that incentives its users to visit five museums and institutions located in Bologna, Italy. The incentive that this app provides amounts to the original artistic content that it unlocks when visiting one of the given locations. Unlike the VersoVerdi app, however, Marsili’s Spirit does not support metaphorical links between museums, which are instead treated as ending points of a treasure hunt.

Now, although the links that the VersoVerdi app unveils constitute an important part of its functionalities, we here emphasize that what really distinguishes such mobile app from any other one is the peculiarity of supporting the emotional browsing of content related to different museums. Emotional browsing is a novel field whose opportunities and implications have been so far only scratched at the surface [3, 4]. In the following we will describe how we have been
3. SYSTEM OVERVIEW

Rather than providing a full description of how our app works from a user’s perspective (an interested user is invited to test our app as soon as it will appear on the Apple and Android stores), we will here, for the sake of conciseness, directly jump to describe the two main features of the VersoVerdi app. The first feature is the emotion recognition mechanism, capable of recognizing whether its user is smiling or not while using the app. Determining the emotional state of a user that is using VersoVerdi is important not only to differentiate the content that is offered, but also to stimulate that sense of discovery that a user can experience as his/her mood changes. The idea has been, hence, that of introducing a sort of “gamification” mechanism inside the app. The second feature is that of linking museums, i.e., exhibit the cultural relations that exist between different institutions. These two features, together, implement the cultural emotional browsing system that is offered by VersoVerdi.

This first mechanism is strongly connected to the second feature, which, instead, amounts to proposing semantic links between different institutions and museums. In practice, following such links, a user discovers multimedia content that follows the mood that one exhibits. We will see how all this has been put to good use in the following.

3.1. Emotion Recognition Mechanism

We designed and implemented an emotion recognition system whose characteristics can be summarized as follows:

- It is capable of capturing and classifying binary emotive states (i.e., happy or unhappy) in less than half a second simply detecting the presence/absence of a smiling person;
- All processing is performed locally, on the smartphone, without requiring any external computational power;
- It is simple and versatile: no complex image analysis techniques are employed and no initial training phase is required when a user first uses it.

To achieve all of these results, we have broken down our problem in a cascade of three sub-problems (Fig. 1). The first problem amounts to individuate the position of the facial features of a user. In practice, when solving this first problem, we adopted a methodology that let us find where the eyes and the mouth of a user are. A second problem, at this point, needs to be solved: find the polygon that best follows the shape of a user’s mouth. Once all of this information is known, we are then capable of addressing the final problem: is the user smiling or not? This last step is solved resorting to a combination of anatomical information and on-the-field experience, as we shall see in brief. Now, let us move on to describe how our mechanism works in its details.

3.1.1. Facial feature recognition

In order to recognize whether a user of the VersoVerdi app is smiling or not, we first of all concentrated on finding the most relevant facial features which could reveal either one or the other of the two given states. Reminding we required a simple and efficient solution, we opted for the following: take into consideration the position of the eyes and the shape of the mouth. Our first problem, hence, has been that of finding the position of these two features.

To individuate the position of each of these elements we utilized the internal CoreImage (i.e., the CoreImage library is specific of the iPhone platforms, we opted for an equivalent solution on the Android version of our app), which allows individuating the positions of the eyes and mouth of a user, under different lighting conditions. Simply speaking, the employed algorithms processes a frame captured by the front webcam and returns the location of the researched features.

Once such step is completed, our mechanism performs the following two operations. The first is that of extracting the area where the mouth lies. The second is to record the distance between the two eyes. Once both of these features are known, the next step is that of modeling the shape of the given user’s mouth.

3.1.3. Modeling a user’s mouth

At this point, our mechanism has obtained the portion of the original frame where the user’s mouth lies. With this assumption, it is possible to find the contour lines that characterize the shape of the given mouth (Fig. 3). In particular, we search for a four-edge polygon, as shown in Fig. 3, whose characteristic is that of following, as closely as possible, the contour of the upper and lower lips.
In order to reach this result (i.e., find the lip following edges) we proceed applying a cascade of filters that: (a) reduce the noise (e.g., lighting changes) that is typically found in the captured image, and, (b) further segment the boundaries of the mouth (i.e., remove any pixels that are not useful, as those that represent parts of nose, beard, etc.). In Fig. 4 we show all the steps that are taken to reach the result of describing the shape of a user’s mouth through the identification of the four points, i.e., the four vertices, which uniquely identify our polygon (Fig. 4), simply analyzing the portion of frame that contains the user’s mouth.

![Fig. 3. Area containing a user’s mouth.](image)

The first operation we perform, hence, is that of reducing noise. This is done utilizing a cascade of three filters. The first filter is a Gaussian one, whose result is that of removing any image snags and artifacts due to jpeg compression. The second filter simply performs a one-channel conversion: the resulting image is a gray-scale one. The final filter that is applied, during noise reduction operations, is a median one. This reduces isolated peaks of intensity, it in fact typically utilized to reduce “salt-and-pepper” noise.

Once noise has been dealt with, it is then possible to focus on the extraction of the four points (i.e., vertices) that identify our sought image. This is done applying a cascade of five filters. The first filter that is applied is a Laplace one. In brief, this filter performs differential operations of the second order (i.e., second order derivatives) on the input pixels, thus highlighting the variations of intensity, which surround a pixel. In essence, the role of this filter is that of extracting the contour of the objects that are represented in an image, enhancing their details.

The resulting image is then fed to a thresholding filter, which converts the input image into a binary one. In the resulting binary image, white pixels represent contours. Unfortunately, however, our resulting binary image may still contain contours that do not represent the mouth (e.g., the beard or the nose of a user might be interfering). For this reason we employ a morphological opening filter, a filter that is capable of suppressing all those shapes that do not follow a predefined one. In practice, if the morphological filter is set with shapes that are similar to the sought ones (i.e., in our case the mouth), it is capable of eliminating all those that greatly differ.

As isolating the contours that describe the mouth from all other ones is the most complex task of our algorithm, we implement another operation to fulfill such scope. In essence, we apply an additional filter that simply: (a) finds the centroid of each contour, (b) computes the distance between the centroid and the center of the mouth, and, (c) eliminates any contour that lies beyond a given distance (such distances vary with the positions, as we take into account what a possible shape of the mouth may be).

We finally search for the vertices that identify our polygon. The vertices that are found during this phase can be essentially of two types: (a) exact (i.e., points that are found with a very high degree of precision and certainty), and, (b) approximate (there points have not been detected with precision, but we assume they fall in a given position, resorting to anatomical considerations). We then proceed approximating the contour that has been so far found inscribing a four-sided polygon, in this way we are able to greatly reduce the computation resources needed to search for the vertices of each contour. In particular, the rightmost and the leftmost points are obtained as the leftmost and the rightmost points of the contours that have been so far individuated. We adopt the same procedure for the remaining two points: the upper and lower lips are detected as the upper and lower points of the contours.

We now have found the four-sided polygon that describes the shape of the mouth. We have now reached the final step: decide whether the given user is smiling/not smiling (i.e., happy/unhappy). To perform such operation we resorted to a large number of tests that revealed us the shapes that may be expected in any of the two cases. In essence, we found that six are the patterns that may typically occur, where three of such patterns can be classified as happy and three as unhappy. The described methodology, which uses simple techniques capable of heavily reducing computational cost, is capable of distinguishing a happy expression from an unhappy one, on average, in less than half a second. We show the six different patterns in Fig. 5. The happy polygons reveal a piece of common knowledge and are very closely resemble how any of us would sketch a smiling person (or also a person who is laughing). The same may be said for the unhappy set of patterns, thus showing that it is possible to build an emotion recognizer utilizing simple geometrical and anatomical considerations.
However, this fact will be further corroborated with the experiments that will be described in Section 5.

![HAPPY vs UNHAPPY](image)

Fig. 5. Classifying a happy user from an unhappy one.

3.2. Linking Museums

The VersoVerdi app has been inspired, during its design, by a metaphor where each museum and institution (i.e., a world of knowledge and culture) is represented as a planet. The planets (i.e., museums) that compose the universe can be browsed simply moving a device around, in the air. In fact, the represented universe covers a wide portion of space which can be simply navigated sliding the screen of a portable device up and down, left and right (e.g., Fig. 6). Depending on the mood that has been detected, the links between museums do not change, but the visible content does. This means that a happy user, for example, will be enabled to explore all the positive content related to (and hence linked to) the museum s/he is interested in, but not the negative (i.e., unhappy, or better, melancholic) content that will hence remain hidden. This means that the content that is presented is effectively driven by the state, the mood, of who is using the app. In Fig. 7 we can appreciate the links that have been created between the different planets in VersoVerdi.

![The VersoVerdi app universe](image)

Fig. 6. The VersoVerdi app universe.

4. PRELIMINARY EXPERIMENTS

We are here offering only a set of preliminary experiments that solely focus on the performance of our emotion recognition mechanism. As we can see from the results shown in Table 1, three different persons, with three different guises, tested the mechanism smiling or remaining serious. We observe that the configuration where the mechanism performed worst is the one experienced by an unhappy goateed user wearing glasses.

![Links connecting planets](image)

Fig. 7. Links connecting planets (i.e., museums).

<table>
<thead>
<tr>
<th>Table 1. Emotion Recognition Mechanism performance.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happy</td>
</tr>
<tr>
<td>------------------------</td>
</tr>
<tr>
<td>Male (glasses, goatee)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Male (no glasses, beard)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Female (glasses)</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

5. CONCLUSION

VersoVerdi is a mobile app celebrating the great composer Giuseppe Verdi. An original aspect is that it supports the emotional browsing of cultural information. To the best of our knowledge this is the first experience of such kind.

6. REFERENCES


