Browsing museum image collections on a multi-touch table

Gianluigi Ciocca [a,*], Paolo Olivo [b,1], Raimondo Schettini [a]

[a] Dipartimento di Informatica Sistemistica e Comunicazione (DISCo) – Università degli Studi di Milano-Bicocca, Milano, Italy
[b] Institut National de Recherche en Informatique et en Automatique (INRIA) – IRCICA, Parc Scientifique de la Haute Borne, Villeneuve d'Ascq, France

ARTICLE INFO

ABSTRACT

In this paper we present a multi-touch tabletop system for browsing image databases, conceived for museums and art gallery exhibitions. The system exploits an innovative image browsing paradigm and image retrieval functionalities to perform natural and intuitive user interaction: users can explore the image database by handling digital pictures by means of touch gestures or by the use of a predetermined set of physical objects; once one of these objects is placed on the table, it is automatically recognized and the associated function is triggered. The set of objects and the function associations can be dynamically configured. An innovative feature of our application is that users can interactively create and manipulate image clusters where images are grouped according to their pictorial similarity. This is achieved by placing one or more specific tangible objects on the table surface. The system has been evaluated on a collection of photos organized in groups according to the UNESCO picture categories. The usability tests, performed considering different user categories, show that users consider the application to be attractive and interesting.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Museums and archives call for methods and tools to navigate their catalogues and to provide facilities for searching, browsing, clustering and visualizing different kinds of visual data and related information. In designing a system able to manage such data, several problems have to be addressed. The most relevant are:

- which navigation paradigms should be offered to users?
- how can non-expert users interact with the system in an intuitive way?
- can these paradigms be appreciated also by experts?

Given the large audience and diversity of the visitors, museums are an interesting target where interactive technologies can be experimented. The kind of information displayed in the exhibitions allows the design and testing of different interaction paradigms, and this explains the interest in developing systems and creative applications that allow visitors, curators and artists to effectively manage, manipulate and explore information. Information exploration may include data browsing or searching for a specific element [1]. In particular, museum visitors often do not have clearly defined goals in mind, but they casually explore repositories with respect to personal interests. For this reason interactive applications must be designed considering how to facilitate free and casual exploration as much as possible, rather than giving them predefined and structured information. Moreover, we must take into account that the possible audience of the museum exhibitions is heterogeneous with diverse educational backgrounds, experience towards technology, and not all people are computer literate. This implies that

* Corresponding author. Tel.: +39 0264487922.
E-mail addresses: ciocca@disco.unimib.it (G. Ciocca),
paolo.olivo@inria.fr (P. Olivo), schettini@disco.unimib.it (R. Schettini).
[1] The work was done while this author was working at DISCo.

the usual interaction via keyboard, mouse and screen is not always the best choice (and is far from being appealing). Applications for public usage must instead encourage interaction (usually using large displays), while hiding the underlying technology as much as possible. The use of interactive applications available to the public also allows the collection of feedback from visitors which can be used to adjust the setting of the current exhibition (add or remove data) or to plan future exhibitions. All these considerations make it very important to define and design the overall system, the visual representation of the information, and to select and develop the appropriate mechanisms to interact with it.

Due to our familiarity with physical tables, digital tabletops displays are one of the most suitable technologies to be used to display information interactively in a museum exhibition. Tabletops are currently used in many museums and galleries [2,3]. The common traits of these installations is that they usually offer a large horizontal display surface as a workspace, mimicking a normal table surface; and users can interact with the tabletops using simple gestures of the hands (usually by finger touching) or physical objects as tools for manipulating the information.

In this work we present a multi-touch tabletop system to be used in the context of museums or art gallery exhibitions for the browsing of image databases by combining an innovative image browsing paradigm and image retrieval functionalities. Images can be browsed freely by category or by interactively created clusters of similar images. The interaction with the displayed images is achieved by directly touching the screen surface or through the use of physical objects that are automatically recognized and classified by their shape. The use of tangible objects for image browsing was first exploited in the Photohelix system [4] where pictures are spatially organized around a tangible object; users can browse the collection by rotating the object. In Photohelix, a mechanism for manually organizing a set of pictures is also supported. Here we exploit tangible objects not only as a means to (passively) browse pictures, but also as an instrument to (actively) organize data via a semi automatic clustering mechanism coupled with a new visualization paradigm. The system exploits a "magnet" metaphor where similar images (from the visual content point of view) are attracted to a target one, while dissimilar images are repulsed. Our browsing application provides an overview of the image collection for the user and allows for intuitive navigation throughout the image database. As stated in [5], the user's cognitive system will play an active role during image browsing. This is an important feature to be exploited in an application aimed at attracting people's attention, and at promoting casual information exploration.

The architecture of our multi-touch tabletop system is shown in Fig. 1. The hardware level corresponds to the physical structure of the table, computer, sensing devices and other support materials. The application Multi-Touch Interface (MTI) is used to get the data from the hardware sensors, detect user interactions, interpret them and send appropriate instructions to the front-end applications.
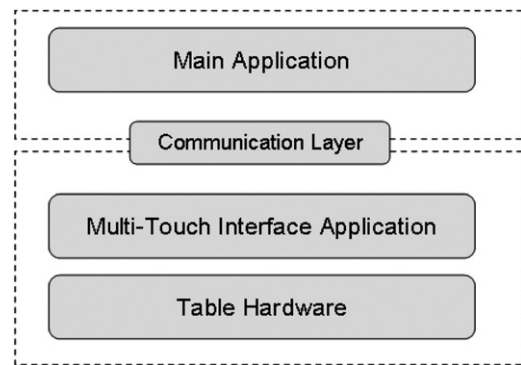

Fig. 1. System's architecture.

The MTI application and the front-end application communicate via a Communication Layer, which exploits TUIO,[2] the widely used communication protocol created by Kaltenbrunner.

2. Related works

One of the first tabletops systems designed to support casual exploration of information is "ThePond" [6]. Other examples of tabletops installation in Museums are the "floating.numbers" by the company ART+COM, developed for the Jewish Museum in Berlin,[3] the "Tree of Life" developed for the Museum of Natural History in Berlin [7] also by ART+COM, the "EMDialog" [8] and the untitled work in [9]. If the underlying application and technology allow it, different visitors can use the tabletop surface at the same time via multi-touch interaction. This has the positive effect of enhancing the museum experience by encouraging the visitors to use the application in a cooperative and collaborative way.

Multi-touch technology is rapidly growing in many different research centres where new features and applications are continuously developed. This technology enables users to interact with applications by means of touch gestures and to handle digital objects by reproducing similar behaviours as if they were real. The multi-touch concepts was first published in 1982 in "A Flexible Machine Interface" [10] by Mehta. The first multi-touch systems designed for human input into a computer system were the "Multi-Touch Tablet" [11] presented in 1985 by Buxton and the "Digital Desk" [12] presented in 1991 by Wellner. More recent developments are the invention of the "FTIR" technology [13] by Han, who also presented the "Multi-Touch Collaboration Wall", and the development of "Surface"[4] by Microsoft in 2007. Today many commercial products (e.g. smartphones, tablet-pc, desktop monitors, etc.) use multi-touch interfaces, and the enhancement of this interaction paradigms is a central topic of the research efforts of big companies like Apple Inc., Microsoft Corporation, Samsung, Asus,

---

Hewlett–Packard, etc. Due to the decreasing cost of the hardware, hobbyists have developed extremely low cost touch screen technologies using web cams, projection of image, and software that senses touch patterns via the web cam. Projects such as the "reacTable" [14], created by the Music Technology Group at Pompeu Fabra University in Barcelona, and the "Community Core Vision",[5] developed by the Natural User Interface group (NUIGroup), are two of the pioneer examples. Open source frameworks have been developed by diverse communities to ease the creation of multi-touch applications (e.g. [15,16]). The availability and easy access to cheap hardware and cheaper software inspired many developer to approach multi-touch technologies, and a great number of projects are being developed every year on increasing rate [17].

A huge competition is currently running in developing applications for browsing and searching large image databases. Very popular social networks like Picasa,[6] Flickr[7] and DeviantArt[8] are used for sharing pictures, which mainly retrieve using tags or textual information; Google Image[9] also exploits pictorial features for a content based image retrieval, but it still presents the output as a grid of images. Current research tries to develop a more dynamic and flexible image visualization, like in Google Swirl[10] or as presented in Nguyen's work [18]. So it is not unexpected that among the first applications developed for multi-touch tabletops or in general for multi-touch displays, the most popular are those for browsing image collections.

3. Table design

The system is intended to be deployed in cultural exhibitions, and then it should be designed to attract people's attention and to be usable in a potentially engaging way. The user interface should enable different users to explore the database at the same time, enabling individual or collaborative usage. To enhance users' experience, we have designed the system so that users can interact with the application with both hand gestures and physical objects to be placed on the table surface. As soon as a physical object is placed on the table, it is automatically recognized, and a virtual object is shown where the real object has been detected. Any movement of the real object is propagated to the corresponding virtual one, and when the real object is removed from the table, also the virtual correspondent is removed. Each object triggers the execution of a predefined action. In the following subsections, we will describe the physical construction of the table and the underlying MTI application responsible for the detection and localization of the users' interactions with the table surface.

3.1. Table setup

We decided to build the table from scratch following the standard guidelines for a DSI (Diffused Surface Illumination) table. The table setting is shown in Fig. 2. The table has been built with a metallic structure 77 cm high with a surface area of 80×60 cm corresponding to an active area of about 70×50 cm. We selected these dimensions for the development of the table prototype. A much larger table can be constructed if necessary. The semi-transparent top surface is constructed from a 10 mm Endlighen acrylic plexyglass that allows the IR light to be uniformly propagated within the surface. A sheet of simple tracing paper, installed under the plexyglass, is used as a projection surface. The plexyglass surface is illuminated by a ribbon composed of 150 infrared LEDs of 880 nm wavelength, uniformly distributed along all four table edges. A Compaq iPAQ MP4800 with a resolution of 1024×768 pixels is used as back projection source. A mirror is used to increase the projector's optical path. This mirror can be removed by using a more expensive short throw projector. The table surface is imaged by a Sony PlayStation Eye camera capable of capturing standard video with a frame rate of 60 Hz at a 640×480 pixel resolution. The camera has been modified to block out visible light while maintaining IR sensibility. Projector and camera are connected to a PC running both the image browsing application and the MTI application. The idea exploited by a DSI tabletop is that finger and object touches on the table surface (i.e. the contact regions) deflect IR light toward the camera in the form of image blobs. The MTI application collects the images from the camera, process them via an image processing pipeline in order to extract the blob locations and geometries. This information is used as interaction inputs for the main browsing application.

Fig. 3 shows the regions where fingers touch the table surface (small blobs) and objects have been placed on the table (large blobs). Clearly, only the contact area of an object with the table surface is visible by the camera. The
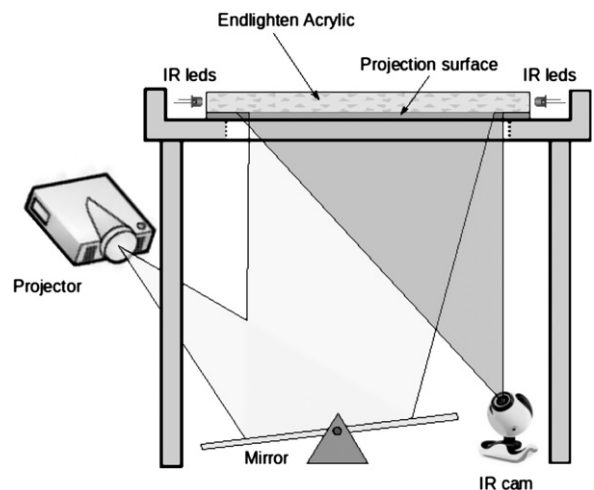
---

[5] http://ccv.nuigroup.com/
[6] http://picasa.google.com
[7] http://www.flickr.com
[8] http://www.deviantart.com
[9] http://www.google.com/imghp?hl=en&tab=wi
[10] http://image-swirl.googlelabs.com

Fig. 2. The table configuration is a DSI setting.

captured image is processed using different image enhancement techniques, and as a result a binary image is produced that shows the contact regions in a more detailed and precise way. Position and geometry of the contact regions are then computed from the binary image and used to classify them as finger touches or physical objects. Objects are further classified based on the shape of their contact area. To allow the recognition of dynamic gestures (i.e. moving a finger or an object on the table surface) the MTI application tracks the contact regions over time. To this end, an identifier is associated to each region and is maintained until the region disappears from the table (i.e the object has been removed or the finger raised).

3.2. The multi-touch interface application

We have developed the Multi-Touch Interface (MTI) application on the basis of the application Community Core Vision (CCV), version 1.2, developed by the NUIGroup. The CCV is an open source and cross-platform application that is widely used in the development of multi-touch interactive tables based on optic technologies. The CCV is able to detect finger touch regions (cursors); and we extended its functionalities for the
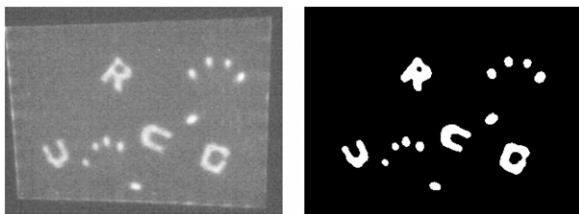

Fig. 3. The finger and object contact regions as seen by the camera (left), and after image processing (right).

detection of object regions. Fig. 4 compares the CCV v1.2 with the MTI application.

The CCV v1.2 package determines only two descriptors for each detected region: the region's centroid that can be used to define the interaction's position on the screen, and the region's area that is used in object recognition for rejecting misleading blobs. To these descriptors we have added another one to be used for physical object classification. The literature offers a huge variety of image region descriptors [19] such as Fourier descriptors, boundary descriptions, topological descriptors, principal components, etc. We have chosen the Hu's invariant moments [20] since they describe the object's shape independently from its position, orientation and scaling factor, and compute quickly.

For a grey scale or binary image with pixel intensities $f(x,y)$, raw image moments m_{pq} are calculated by

$$m_{pq} = \sum_x \sum_y x^p y^q f(x,y) \tag{1}$$

In order to get translational invariant descriptors, central moments μ_{pq} have been defined as:

$$\mu_{pq} = \sum_x \sum_y (x-\overline{x})^p (y-\overline{y})^q f(x,y) \tag{2}$$

where

$$\overline{x} = \frac{m_{10}}{m_{00}}, \quad \overline{y} = \frac{m_{01}}{m_{00}} \quad p,q = 0,1,2,\ldots,+\infty$$

Point $(\overline{x},\overline{y})$ is the image centroid and is calculated using moments' equation (1). Moments η_{pq}, where $p+q \geq 2$, can be constructed to be invariant to both translation and changes in scale, by dividing the corresponding central moment by the properly scaled moment μ_{00}, using the following formula:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}}, \quad \gamma = \frac{p+q}{2} + 1.$$
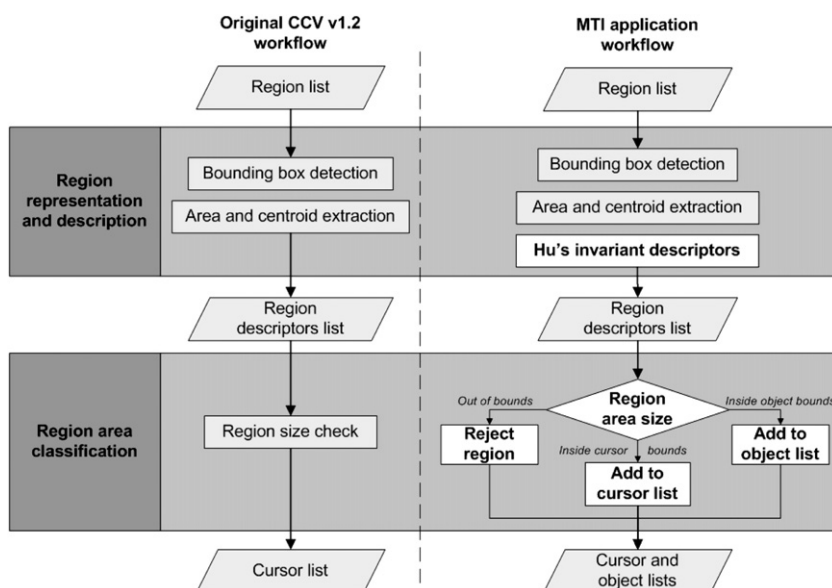

Fig. 4. The CCV v1.2 workflow vs. the MTI application workflow.

The set of Hu's invariant moments is defined as:

$$\phi_1 = \eta_{20} + \eta_{02}$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$

$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + 3\eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2]$$
$$+ (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$

$$\phi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2$$
$$+ 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})]$$

$$\phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} - \eta_{03})^2]$$
$$+ (\eta_{30} - 3\eta_{21})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (3)$$

These moments are invariant to translation, rotation and scale changes. The first moment, ϕ_1, is analogous to the moment of inertia around the image's centroid, where the pixels' intensities are analogous to physical density. The last one, ϕ_7, is skew invariant, which enables it to distinguish mirror images of otherwise identical images. We compute the Hu invariant moments on the object regions of the binarized image (Fig. 3, left). The shape of each detected object is thus described in terms of the feature vector $[\phi_1, \phi_2, \ldots, \phi_7]$.

The region classification shown in Fig. 4 is based on the area of the detected regions and is basically aimed at distinguishing between cursor, object and noise regions. Object's shape classification can be performed either by the MTI application or by the main application. Although it requires the inclusion of a classification algorithm within the main application, the latter is more flexible since it allows different classification strategies to be used, depending on the application requirements.

Object classification involves techniques to learn (or build) a knowledge model about object classes by a training process. In our application the training set has been acquired by placing a few objects (one at time) on the table surface and letting the MTI application detect the objects and determine their invariant moments. Each acquisition represents an object's model so a single object can be represented by several models. Having more than one model for each object makes the classification both robust and simple by using a k NN (k-Nearest Neighbour) classifier [21].

The k-Nearest Neighbour classifier is a non-parametric classification algorithm, where an instance of a data point x is compared with the set of k labelled points (models) closest to it, and they vote for the label assigned to the point (i.e. to x is assigned the label belonging to the majority of the k models). Formally, given a set N of labelled data points, the set A_x of the k-Nearest Neighbour points of a data point x is given by:

$$A_x = \{y : \|x - y\| \le \|y - q\|, q \in N - A_x\}$$

$$|A_x| = k \quad (4)$$

where $\| \cdot \|$ is a suitable distance measure. Let *label* be a function that maps a data point to its class label, then the set of points in A_x having label i with $i \in \{1, \ldots, L\}$ can be defined as:

$$C_i = \{y \in A_x : label(y) = i\} \quad (5)$$

A point x is assigned a label c if:

$$|C_c| > |C_j| \quad \forall j = 1, \ldots, L \text{ and } j \ne c \quad (6)$$

In our application we set the classifier with $k = 3$, and two feature vectors are compared using the Euclidean distance. Moreover, to make the classification stage more robust we have included a rejection class for "unknown objects". To classify the objects on the table surface, fiducial marks could be used [22]. We chose to rely on general purpose shape descriptors because fiducial marks must be physically attached to the objects, coded and decoded accordingly. Moreover, shape descriptors can easily and transparently be used to recognize fiducial marks with characteristic shapes as well. Fig. 5 shows some examples of physical objects used in our tests.

After touches and objects have been detected and classified, they are sent to the browsing application using the TUIO 1.0 protocol [23]. The TUIO protocol is an extensible formal structure for defining and transmitting user interactions via TCP or UDP protocols. It was developed within reacTIVision [22], and adopted by several other projects related to tangible and multi-touch interaction. The TUIO protocol enables the MTI application and the browsing application to communicate following a client–server architecture; thus each application needs to implement a TUIO sender and listener module respectively. The TUIO protocol defines a standard set of profiles for tracking objects and cursors on two dimensional surfaces. Since the standard profiles do not support the object's shape information, the TUIO protocol has been extended adding a new profile that enables the transmission of the Hu seven invariant moments. This extension required the modification of both the TUIO sender and the listener modules as well.

Fig. 5. Examples of tangible objects used in our application. On the left, objects made of cardboard and parcel tape. In the middle, objects made of cylindrical Plexiglass with the shapes printed and attached to the base and an explanatory image attached to the top. On the right, objects made of velvet.

4. Application design

Our tabletop application is a part of a larger project developed in collaboration with the Lombardy Region and financed by the Cariplo Foundation that aims to preserve and disseminate the local cultural traditions of the Lombardy Region. The main objective is to enable users to browse collections of photos organized in groups according to the UNESCO picture categories [24]: arts, orality, nature, knowledge, and rites. Images are displayed on the table surface through *virtual pictures*. Information associated to each image (e.g. authors, categories, descriptions, etc.) is stored in a database and may be retrieved and visualized by the virtual pictures on request. A user can simply browse the predefined image categories or create groups of similar images via an interactive image clustering algorithm. The latter is achieved by exploiting a magnet metaphor applied to the images. Users can interact with the virtual pictures by touches and objects. The use of tangible objects allows the whole table surface to be usable for image display since it does not require space to be reserved for fixed graphical interface elements such as buttons and menus. Moreover, we think that by using objects that casual users can grab and drag around the surface will make the interaction with the application more enjoyable and "real-feeling" and less abstract. Specifically, in our application we have defined three basic actions that can be performed by physical objects: (a) browse the image in the database by importing them on the table via the *importer object*; (b) cluster similar images via the *magnetizer object*; (c) remove images from the table via the *remover object*.

The application has been developed using the Java Monkey Engine (JME) graphical framework.[11] JME is an open source, high performance, 3D scenegraph based graphics API, written in Java. This choice is justified by the wide support that JME offers for managing 2D and 3D object manipulations (i.e. translations, rotation and resizing), picking and collision, and for the included physics engine library. Moreover, JME allows us to develop cross-platform applications.

4.1. Virtual pictures

The most important visual element in our application is the *virtual picture* (Fig. 6): a manipulable, digital representation of a photo.

Virtual pictures can be moved, rotated, resized and removed from the table. All these actions can be performed using finger touches (see Fig. 7). For example, a picture movement is achieved by touching a virtual picture and then dragging it on the table while maintaining the touch. Rotation is performed using two finger touches: the first touch defines the pivot location, while the second touch is used to move the virtual picture around the pivot point. Resizing uses the "pinch" paradigm: moving apart two fingers on the virtual picture
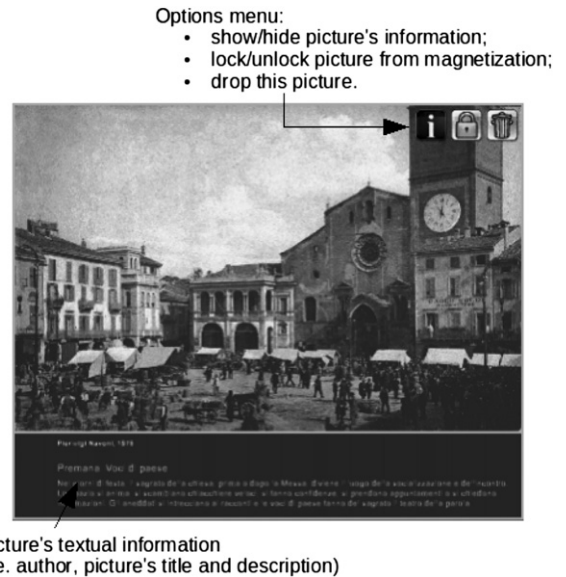

Fig. 6. The virtual picture. On the top right the options menu can be seen with the set of icons that can be used to manipulate the virtual picture: (from left to right) display picture info, picture lock and picture removal.
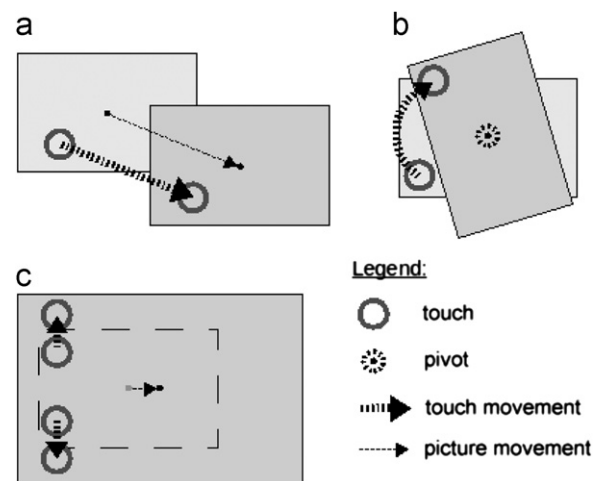

Fig. 7. The basic actions that can be performed on a virtual picture using finger touches: (a) movement, (b) rotation, and c) resizing.

---

enlarges the picture; moving them near each other shrinks the picture.

These basic actions are similar to the ones widely used in other touch devices such as the mobile phone or the tablet. Other actions are available through the option icons that are associated to each picture. These icons make it possible to show or hide the textual information associated to the picture, lock or unlock the picture, and remove the picture from the table surface. The first and last actions are self-explanatory. The lock icon is used to inhibit certain actions on the image itself. For example, a locked image cannot participate in the image clustering process: its position on the table is not changed even if it is under the effect of the magnetizer objects. A locked

---

picture can be removed from the table surface only by using the remove icon since the remover physical object has no influence on it. During the interaction, the locking mechanism is useful when a user wants to "protect" one or more pictures from being captured by other users via their magnetizer objects. Similarly, on a picture-cluttered table, a picture can be locked while safely removing all the other ones from the table. Locked pictures have a semi-transparent padlock icon superimposed on the image to mark their status.

4.2. Image import and free browsing

An image in order to be displayed on the table surface must be imported from the database. This operation is achieved by placing a physical object associated to an importer object on the table surface (Fig. 8).

As soon as a picture importer is placed on the table, a few previews of the images belonging to that category are displayed around it. The images are displayed with different sizes hinting that other images are behind the smaller ones. When the real object is rotated, the image previews are scrolled around it, so that it is possible to browse all the images in the current category. Users can import any of these images by dragging them on the table, or place all the images at random positions by using the button on the left side of the object. An importer object can load pictures from a single image category (arts, orality, nature, knowledge, or rites); because of this, in order to browse the entire database, at least five real objects must be enroled as importer objects.

As can be seen, all the images around the importer object are displayed with a preferred, fixed orientation.

This may seem a strong constraint on how the table surface is used. In our application physical objects have no "up" or "down" orientations so we do not have a reference orientation to be used to display the images accordingly. One way to cope with this problem would be to display the images in a circular pattern with their "lower" edge facing the physical object. This would have allowed different users to use the table from different sides, but at the same time would have resulted in few images being fully readable. Our choice makes them all readable, and if the table surface is large enough, several users can approach the table from the same side without being greatly impeded in their actions.

Pictures can be removed from the table surface either by using the remove icon in the picture's options menu or by using the remover object. The remover object (see Fig. 9), which is activated by a corresponding physical object, has an area of influence (displayed as a coloured circle) that can be resized by simply rotating the object. Any picture whose centre falls inside the area of influence of the remover object is sucked out of the table surface. Virtually, the remover object works like a black hole. Users can also drag selected pictures within the area of influence. As stated before, locked pictures cannot be discarded using the remover object until they are unlocked.

4.3. Image browsing via interactive clustering

One innovative feature of our application is image clustering. Users can create groups of similar images by exploiting a "magnet" metaphor: a user can magnetize an image by placing a magnetizer object near it. When an
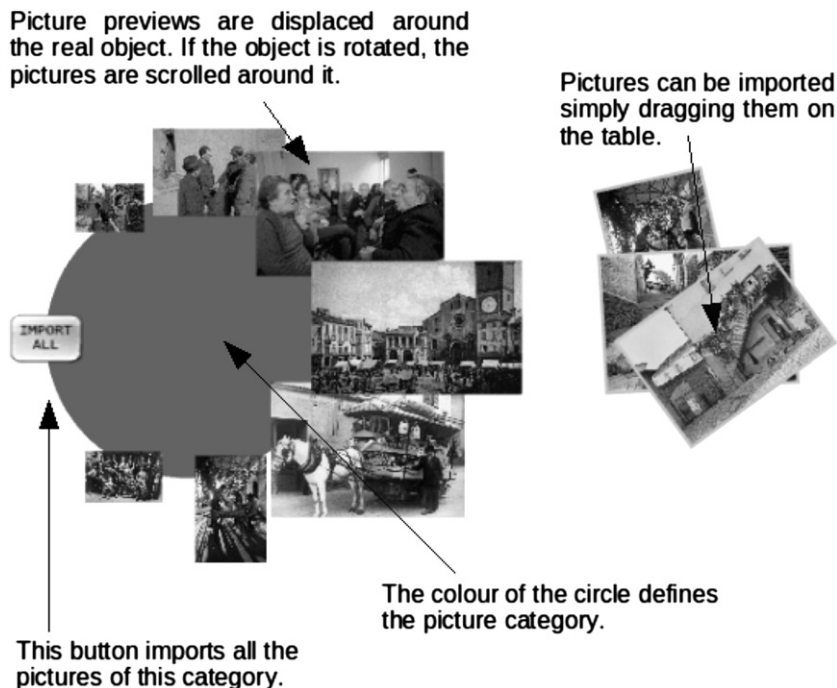

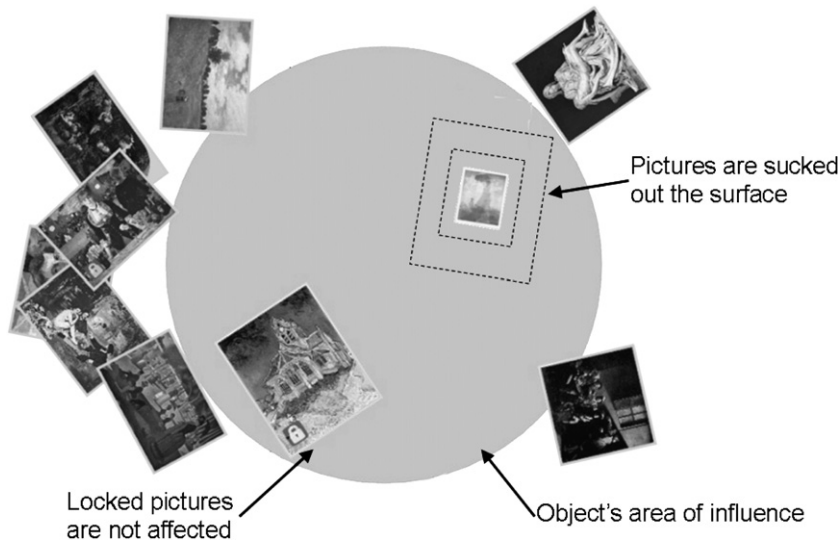**Fig. 8.** The picture importer virtual object.

Fig. 9. The picture remover virtual object.

image is magnetized, all the images that are similar to the magnetized one are attracted near it. The attracted images are moved around following a spiral path where the magnetized image is at the initial position, while the others images are placed in sequence along the spiral circles at regular intervals, ordered by decreasing similarity: the more similar images are placed along the inner circles, while the less similar ones are placed along the outer circles. The image size is also affected: inner images are bigger than the outer ones.

More than one magnetizer object can be placed on the table at the same time to create different clusters of similar images. All the images on the table surface rearrange themselves near their most similar magnetic companion. Each magnetic object has an adjustable area of influence, which can be controlled by varying the orientation of the associated physical object. Any image situated in this area is captured by the object's magnetic field. The magnetic field can also be enlarged to allow users to cooperate in creating image clusters by sharing all their images. For example, if an image is placed in the area where two or more magnetic fields overlap, it is attracted to the most similar magnetic image. To split all the images on the table in different groups, the magnetic fields can be enlarged to encompass the whole table surface. A user can also lock an image in order prevent it from being subjected to the influence of any magnetic field. A scrollable menu displayed on the left of the magnetizer object can be used to change the status of the attracted images (e.g. lock) or their visualization properties (e.g. picture spacing, spirals density, sizes, etc.). An example of the magnetization procedure is shown in Fig. 10.

4.3.1. Image description and clustering procedure

Images are similar if they share a similar pictorial contents, whatever the category they belong to. In order to describe the pictorial contents we have used some image descriptors taken from the content-based image retrieval field [25]. They describe elementary visual characteristics such as shape, colour, orientation and texture of the objects contained in the images. Each descriptor is defined by a set of numerical values; therefore it is possible to use a distance function to define the similarity between images based on the descriptors. Image descriptors are computed on each image during the database population and stored along with the image information. Descriptors are then automatically retrieved when the images are imported on the table to be used during the clustering process. Many different image descriptors exist in the literature, and for practical purposes more than one descriptor is usually exploited. In our application we have chosen three image descriptors: one that determines the colour content of the image and two taking into account the structure of the image regions. Specifically, the image descriptors used here are a subset taken from [26]: Colour Coherence Vector, Colour Regions Composition, and Wavelet Statistics.

The Colour Coherence Vector is an image descriptor first introduced in [27]. Given a colour quantized image, the i-th entry in the Colour Coherence Vector is a pair of values (α_i, β_i) where α_i represents the fraction of pixels of colour index i belonging to regions of size greater than a threshold (coherent pixels), and β_i represents the fraction of pixels of colour index i belonging to regions of small size (incoherent pixels). The distance (D_1) between two Colour Coherence Vectors V^1 and V^2 is computed using the following formula:

$$D_1(V^1, V^2) = \sum_{j=1}^{n} |(\alpha_j^1 - \alpha_j^2) + (\beta_j^1 - \beta_j^2)| \qquad (7)$$

The Colour Regions Composition feature captures how the colour regions are spatially distributed within the image. The feature is composed of four values: fragmentation, dispersion, horizontal symmetry, and vertical
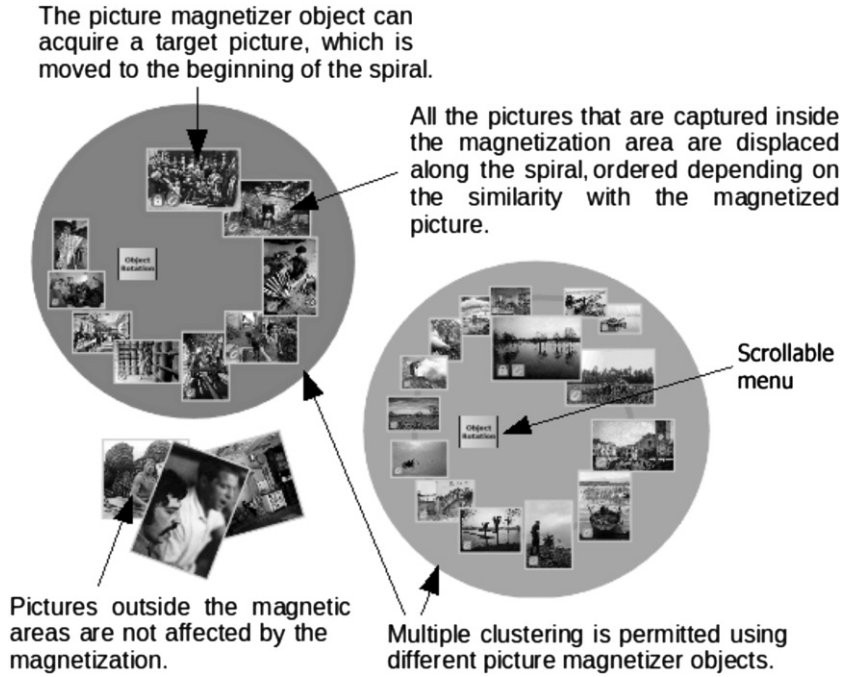
The picture magnetizer object can acquire a target picture, which is moved to the beginning of the spiral.

All the pictures that are captured inside the magnetization area are displaced along the spiral, ordered depending on the similarity with the magnetized picture.

Scrollable menu

Pictures outside the magnetic areas are not affected by the magnetization.

Multiple clustering is permitted using different picture magnetizer objects.

Fig. 10. The interactive clustering activity.

symmetry. The distance (D_2) between two colour regions composition feature vectors C^1 and C^2 is computed by using the L_1 distance norm.

Wavelet analysis provides representations of the image data in which both spatial and frequency information are present. Wavelets are computed by applying a bank of filters on the image, and four new images (bands) are obtained (one is the image sub-sampled). Multi-resolution analysis is achieved by reapplying the bank of filters on the sub-sampled image. In this application Daubechies' wavelet decomposition is applied three times on the luminance image to generate ten bands. Two energy values, the mean and standard deviation, are computed on each band for a total of 20 values. The distance (D_3) between two wavelet statistics feature vectors W^1 and W^2 is computed using the Euclidean distance.

An image I is thus described with the tuple $\langle V, C, W \rangle$. To determine the distance between two images I^1 and I^2, their descriptors are independently compared, and the three distances obtained are combined together in a final similarity value S:

$$S(I^1, I^2) = \phi(D_1(V^1, V^2)) + \phi(D_2(C^1, C^2)) + \phi(D_3(W^1, W^2)) \quad (8)$$

where ϕ is a function used to transform the distances in similarities in the range [0,1].

The clustering algorithm used in our application is very simple and can be considered as a k-means clustering algorithm [21] without cluster updates. In our application, the magnetized images represent the initial cluster seeds, and only one step of the clustering algorithm is performed. After all the magnetizable images (i.e. those that are not locked or are outside the influence of the magnetic fields) have been assigned to the nearest seed,

they are sorted according to their similarities to the respective magnetic image, and then each group of images is displayed.

5. System evaluation

In order to measure how well the users interact with the application and how easily certain tasks can be achieved, we performed a usability test involving several users. Subjects were selected taking into account a broad range of ages, expertise and educational backgrounds covering the kind of audience of a typical museum exhibition. Specifically, 16 subjects (Italian native speakers) were enlisted and divided into four groups: average users, expert users, younger users and elderly users. Each group was composed of four subjects, and both males and females were included. In the average user group there were subjects from 14 to 55 years old, with none or very basic knowledge in computers and related applications. The expert user group was formed by subjects from 14 to 55 years old with extensive experience in the usage of computers and multi-touch applications. Younger users were children from 6 to 14 years old, while the elderly users were over 55 years old. For these two latter groups, no assumptions were made on their attitudes towards advanced technology. The number of subjects used in our experiment is in accordance with the guidelines proposed in [28].

The experiment was conducted within our department in an isolated room and consisted in a series of tasks to be performed. The subjects were individually taken into the room and were encouraged to freely use the virtual table and explore the application functionalities. After this brief

adaptation phase we asked the subjects to complete the following tasks:

1. select and read aloud the textual information associated to two images;
2. arrange five images close together without gaps;
3. fill the table surface with images, lock two images at will and remove the remaining images;
4. using the magnetizer objects, create two image clusters: one of images depicting fire and one of images depicting water.

The tasks were devised to test a broad range of the application functionalities and were administered in increasing order of complexity. The first task covered the basic functionalities of the application and required the users to browse an image collection and to locate and activate the info icon in the options menu that showed the textual information associated to the images. The aim of the second task was to test how easily users were able to spatially arrange a subset of images without gaps by selecting, resizing and rotating them. Most of these operations required interaction with multiple touches at the same time. The third task involved the use of the importer physical objects to fill the table with images, to locate and select the lock icon in the options menu and to use the remover physical object. With this task we tested the user behaviours while using different physical and virtual objects at the same time. The third task tested the use of the magnetizer objects and their interaction in creating two clusters of images depicting different concepts. This last task covered the more advanced functionalities of the application.

In order to analyse user behaviours we filmed the subjects throughout the entire experiment; a video camera was positioned to film the entire table surface. We also recorded the user feedbacks following the "think-aloud" protocol [29]: the subjects were asked to describe aloud the operations they were performing and the problems encountered while completing the tasks.

Finally, the subjects judged their experience in using the virtual table and the application by filling in a questionnaire. The questionnaire was subdivided into two sections. The first section was based on the System Usability Scale (SUS) questionnaire developed by Brooke at DEC (Digital Equipment Corporation) [30]. It is composed of a series of statements related to different aspects of the experience, and the subjects were asked to express their agreement or disagreement on the given statement with a score taken from a Likert scale of five numerical values: one expressing strong disagreement with the statement, five expressing strong agreement and five expressing a neutral answer. Fig. 11 shows the SUS section of the questionnaire adapted for our application.

It must be noted that in order to be understandable by all the subjects, the word "table" was used to indicate the whole system, and that the questionnaire was administered after being translated into Italian. Moreover, the questions marked with odd numbers cover positive aspects or feelings in the use of the system; and thus, the higher the score, the better the evaluation is. On the contrary, the questions marked with even numbers analyse negative aspects of the systems; and thus, the lower the score, the better the evaluation is. The SUS uses a "quick and dirty" approach in evaluating the overall subjective usability of a system. A SUS score is mapped in the numerical range 0–100, but it can also be mapped to a different scale as suggested by Bangor et al. [31]. The authors devised three mappings from SUS scores: acceptability ranges, grade scores and adjective ratings, as shown in Fig. 12.

The second section was used by the subjects to freely write comments, suggestions and notes about their experience with the application and the virtual table. The data in this section would be difficult to analyse given its unstructured format, but it was deemed very important since if would allow us to collect insightful observations on how to improve the application by adding new functionalities and by changing or removing existing ones.

6. Discussion and results

Table 1 shows how the different users managed to complete the given tasks. The users were allowed to complete the task at their leisure, no time limit was imposed. They were helped if technical difficulties arose, but to complete the task they were left to themselves. Very few exceptions were made: hints were given when they were stuck with some actions, and only after they were advised to try different paths.

A "Yes" in a cell means the user was able to successfully complete the given task; "No" means that the user was unable to perform the required task. An asterisk near the "Yes" mark means that the user was able to complete the task, but by following a different path of actions from the intended one. Specifically, in Task 3, the users removed the images from the table one a time by using the delete icon button in the options menu instead of the remover object. This does not mean that the user "cheated" the system but that he/she was more comfortable in using a different approach which was also available.

The users in the expert group were able to finish all the required tasks without any problems. We expected this since these users knew touch interfaces and were computer literate. Only one user (EX1) was not able to complete the more complex task, since to complete Task 4 it was also required to interact with the contextual menu of the magnetizer object. The user found it very difficult to read the text on the menu and thus was unable to perform the required actions. The main reason for this problem is the lack of enough free space between the object and the images around it that forced the design of the menu in a rather small region. Probably the non-optimal resolution of the projected screen also contributed negatively to the readability of the texts on the visual interface. We will see that the same issue was raised by other users as well.

The results of the average users are similar to those of the expert users. No relevant problems emerged during the experimentation. Only one user was not able to

System Usability Scale

© Digital Equipment Corporation, 1986.

	Strongly disagree				Strongly agree
1. I think that I would like to use this table frequently	1	2	3	4	5
2. I found the table unnecessarily complex	1	2	3	4	5
3. I thought the table was easy to use	1	2	3	4	5
4. I think that I would need the support of a technical person to be able to use this system	1	2	3	4	5
5. I found the various functions in this table were well integrated	1	2	3	4	5
6. I thought there was too much inconsistency in this system	1	2	3	4	5
7. I would imagine that most people would learn to use the table very quickly	1	2	3	4	5
8. I found the table very cumbersome to use	1	2	3	4	5
9. I felt very confident using the table	1	2	3	4	5
10. I needed to learn a lot of things before I could get going with this table	1	2	3	4	5

Fig. 11. The SUS section of the questionnaire administered to the users after completing the required tasks on the virtual table.
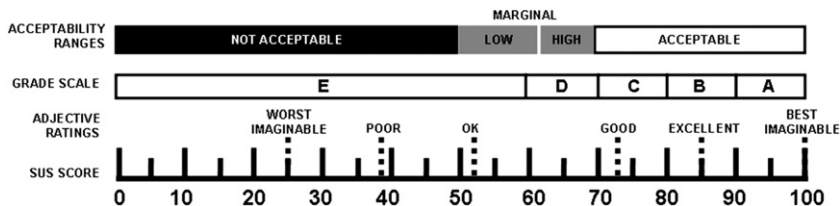

Fig. 12. Adjective ratings, acceptability scores, and school grading scales, in relation to the SUS scores. (Image reproduced from [31].)

correctly interact with the contextual menu of the images. In order to view the text associated with the images, the image must be selected with one touch. This will bring out the contextual menu of the image. While maintaining the image selected with one touch, the action (e.g. info) on the contextual menu must be selected with another touch. If the first touch is released, the contextual menu disappears. The user did not fully understand this behaviour and insisted in using a single finger to interact with the image and thus was not able to maintain the contextual menu visible and select the required action.

The young users were the most excited in being part of an experiment with a multi-touch virtual table application and were not intimidated by the setting. Moreover, they played with the table a lot before performing the tasks. After a while, the more demanding task (Task 4) was abandoned by all the young users. This is probably due to the many steps required to complete it. They

Table 1
Success in completion of the four tasks.

Group	User ID	Task 1	Task 2	Task 3	Task 4
Expert users	X1	Yes	Yes	Yes*	No
	X2	Yes	Yes	Yes*	Yes
	X3	Yes	Yes	Yes	Yes
	X4	Yes	Yes	Yes	Yes
Average users	A1	No	Yes	Yes*	No
	A2	Yes	Yes	Yes*	Yes
	A3	Yes	Yes	Yes	Yes
	A4	Yes	Yes	Yes	Yes
Young users	Y1	Yes	Yes	Yes*	No
	Y2	No	Yes	Yes	No
	Y3	Yes	Yes	Yes	No
	Y4	Yes	Yes	Yes*	No
Elderly users	E1	No	Yes	No	No
	E2	Yes	Yes	No	No
	E3	Yes	Yes	No	No
	E4	Yes	Yes	No	No

Table 2
Agreement votes about positive aspects of the system (higher better).

SUS S	1	2	3	4	5
01	0 (0%)	0 (0%)	4 (25%)	8 (50%)	4 (25%)
03	0 (0%)	2 (12.5%)	2 (12.5%)	8 (50%)	8 (25%)
05	0 (0%)	2 (12.5%)	5 (31.25%)	7 (43.75%)	4 (12.5%)
07	0 (0%)	2 (12.5%)	6 (37.5%)	6 (37.5%)	4 (12.5%)
09	1 (6.25%)	0 (0%)	2 (12.5%)	2 (12.5%)	6 (37.5%)

Table 3
Agreement votes about negative aspects of the system (lower better).

SUS S	1	2	3	4	5
02	7 (43.75%)	7 (43.75%)	1 (6.25%)	0 (0%)	1 (6.25%)
04	7 (43.75%)	3 (18.75%)	4 (25%)	2 (12.5%)	0 (0%)
06	9 (56.25%)	7 (43.75%)	0 (0%)	0 (0%)	0 (0%)
08	10 (62.50%)	4 (25%)	1 (6.25%)	1 (6.25%)	0 (0%)
10	7 (43.75%)	5 (31.25%)	2 (12.5%)	2 (12.5%)	0 (0%)

preferred to freely explore and play with the images on the table.

It can be seen that the elderly users were not able to successfully complete the more complex tasks. The main reason for the failures in Task 3 is the use of the remover physical object. Its role and usage was not fully understood by the users, and thus they ignored it entirely. They found the image lock mechanism difficult to understand and activate. All these users considered it very difficult to read the contextual menu of the magnetizer object. Overall, the elderly users were more comfortable in interacting with the application with their fingers only without the use of other objects (apart from the importer object).

Tables 2 and 3 show the results of the questionnaire. The answers have been divided into two groups: those related to the positive aspects of the application (Table 2) and those related to the negative aspects (Table 3). In Table 2 we can see that the users were satisfied with the usability of the table and the application, with most of the votes being between 3 and 5. Six users did not consider the interaction with the application as easy as they had expected (questions 3, 5 and 7). Only one user (in the elderly group) did not feel very confident in using the table (S09), he was unsure about the purpose of the application and having had no previous experience in the use of computer and touch applications, was very apprehensive while approaching the tasks. Table 3 follows

a similar pattern. Most of the users disagreed with the (negative) statements. Again only one user judged the application too difficult to use; five criticized the complexity of the system and suggested having an operator available to actively help the users while using the application. On the other hand, as it can be seen from the votes statement S09 received, most of the users managed to cope with some difficulties by themselves.

Taking into account all the 10 questionnaire statements, the overall SUS score for our application, computed as described in [30], is 73.97. By comparing the system's SUS score with the other evaluation scales of 12, we can see that our application is within the "Acceptable" range with respect to the acceptability range and is evaluated "Good" using the adjective ratings.

The videos taken during the experimental sessions revealed that our application suffers from a number of issues that should be addressed before deploying the application in the real setting. One of the most frequent complaints expressed by the users referred to the contextual menu of the magnetizer object. As stated before, all the elderly users found it difficult to read the menu; the other users coped with this problem but stated that the text needed to be resized. In order to resize the text, we need to redesign the contextual menu either by eliminating seldom used functionalities or by finding another position where it can be visualized in a larger region. Also the font used in the menu must be substituted with a more legible one. Another issues raised by most users (12 out of 16) was that only after using the application for a time they discovered that by rotating the importer object they can browse the image collection. Initially they thought that the images in the collection were only those few displayed. This was somewhat as a surprise for us since we designed the interface to display the images using different sizes: a large image at the right of the object and then smaller images around the object as they move away from the larger one. We wrongly thought that this visualization modality was a sufficient hint for the users. The same amount of users did not fully understand how to use the remover object, either they did not recognize it among the physical objects or they preferred to use a more traditional "click and delete" approach to remove the images. We need to modify the remover object by adding more visual cues on its usage. Seven users out of 16 found it the image locking concept difficult to understand: they did not understand why a locked image could be manipulated like a normal one (or so they thought). This was partially true since the only two operations not allowed on a locked image were its removal and its change of cluster under the effect of another magnetizer object. We realize that these two

operations have a slightly different semantics and thus must be addressed differently to be more understandable. The term "lock" may be misleading for the user and should be reconsidered. Five users out of 16 had difficulty in resizing the image. They tried to resized the images using the same approach they were familiar with while using a personal computer with a windows based graphical interface, that is by clicking and dragging the lower right corner of the image. After some trials all these users found out how to resize the images using two fingers. Finally, the elderly users were ill at ease in performing actions with multiple touches on the same virtual object (e.g. image rotation). They preferred interaction via single touches. This fact should be investigated more thoroughly in order to design a simplified interface for specific groups of users.

7. Conclusions

We presented a multi-touch table-top system for browsing large picture databases. The system has been conceived to support museums' image archives exploration, but it can be easily upgraded to fit a more wide range of targets. Since expected users will not necessarily be computer experts, we designed interaction paradigms to be as intuitive as possible. Thus, the designed system enables users to browse the database with both hand gestural interaction and by means of physical objects to be placed on the table. In addition to the object's shape, the object's orientation is also used in supporting specific actions. An original features of our application is that users can interactively group similar images by means of automatic clustering functions. We think that expert and non-expert users can benefit from the implemented clustering technique, since it strongly helps image search and allows users to create personal collections, which can be, for example, purchased or printed. The usability test, performed considering different user categories, showed that the most users considered the application to be attractive and interesting. By using the SUS evaluation system, the application received an overall "good" score. The test also helped in pointing out that some issues must be corrected in order to improve its usability for the elderly users. We plan to cope with these issues in the future and perform more field tests, particularly in the collaborative scenario. Additional materials can be found at: http://www.ivl.disco.unimib.it/activities/vtable.

References

[1] C. Müller-Tomfelde, P. Isenberg, U. Hinrichs, M. Hancock, S. Carpendale, Digital tables for collaborative information exploration, in: C. Müller-Tomfelde (Ed.), Tabletops – Horizontal Interactive Displays, Human–Computer Interaction Series, Springer, London, 2010, pp. 387–405.
[2] T. Geller, Interactive tabletop exhibits in museums and galleries, IEEE Computer Graphics and Applications 26 (2006) 6–11.
[3] T. Naemura, Y. Kakehi, T. Hashida, Y. ah Seong, D. Akatsuka, T. Wada, T. Nariya, T. Nakashima, R. Oshima, T. Kuno, Mixed reality technologies for museum experience, in: Proceedings of the 9th ACM SIGGRAPH Conference on Virtual-Reality Continuum and its Applications in Industry, VRCAI '10, ACM, New York, NY, USA, 2010, pp. 17–20.
[4] O. Hilliges, D. Baur, A. Butz, Photohelix: browsing, sorting and sharing digital photo collections, in: Second Annual IEEE International Workshop on Horizontal Interactive Human–Computer Systems, TABLETOP '07, 2007, pp. 87–94.
[5] W. Plant, G. Schaefer, Visualisation and browsing of image databases, in: W. Lin, D. Tao, J. Kacprzyk, Z. Li, E. Izquierdo, H. Wang (Eds.), Multimedia Analysis, Processing and Communications, Studies in Computational Intelligence, vol. 346, Springer, Berlin/Heidelberg, 2011, pp. 3–57.
[6] O. Ståhl, A. Wallberg, J. Söderberg, J. Humble, L.E. Fahlén, A. Bullock, J. Lundberg, Information exploration using the pond, in: Proceedings of the 4th International Conference on Collaborative Virtual Environments, CVE '02, ACM, New York, NY, USA, 2002, pp. 72–79.
[7] E. Hornecker, I do not understand it either, but it is cool - visitor interactions with a multi-touch table in a museum, in: Third IEEE International Workshop on Horizontal Interactive Human Computer Systems, TABLETOP 2008, 2008, pp. 113–120.
[8] U. Hinrichs, H. Schmidt, S. Carpendale, Emdialog: bringing information visualization into the museum, IEEE Transactions on Visualization and Computer Graphics 14 (2008) 1181–1188.
[9] N. Correia, T. Mota, R. Nóbrega, L. Silva, A. Almeida, A multi-touch tabletop for robust multimedia interaction in museums, in: ACM International Conference on Interactive Tabletops and Surfaces, ITS '10, ACM, New York, NY, USA, 2010, pp. 117–120.
[10] N. Mehta, A Flexible Machine Interface, Master's Thesis, Department of Electrical Engineering, University of Toronto, 1982.
[11] S.K. Lee, W. Buxton, K.C. Smith, A multi-touch three dimensional touch-sensitive tablet, in: Proceedings of CHI '85, ACM, 1985, pp. 21–25.
[12] P. Wellner, The digitaldesk calculator: tangible manipulation on a desk top display, in: Proceedings of the 4th Annual ACM Symposium on User Interface Software and Technology, UIST '91, ACM, New York, NY, USA, 1991, pp. 27–33.
[13] J.Y. Han, Low-cost multi-touch sensing through frustrated total internal reflection, in: Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology, UIST '05, ACM, New York, NY, USA, 2005, pp. 115–118.
[14] S. Jordà, G. Geiger, M. Alonso, M. Kaltenbrunner, The reactable: exploring the synergy between live music performance and tabletop tangible interfaces, in: Proceedings of the 1st International Conference on Tangible and Embedded Interaction, TEI '07, ACM, New York, NY, USA, 2007, pp. 139–146.
[15] T.E. Hansen, J.P. Hourcade, M. Virbel, S. Patali, T. Serra, PyMT: a post-wimp multi-touch user interface toolkit, in: Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces, ITS '09, ACM, New York, NY, USA, 2009, pp. 17–24.
[16] U. Laufs, C. Ruff, J. Zibuschka, in: Engineering Patterns for Multi-Touch Interfaces, Workshop of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems, 2010, pp. 52–57.
[17] J. Schöning, A. Krüger, P. Olivier, Multi-touch is dead, long live multi-touch, in: CHI 2009 Workshop: Multitouch and Surface Computing, 2009.
[18] G.P. Nguyen, M. Worring, Interactive access to large image collections using similarity-based visualization, Journal of Visual Languages and Computing 19 (2008) 203–224.
[19] D. Zhang, G. Lu, Review of shape representation and description techniques, Pattern Recognition 37 (2004) 1–19.
[20] M.-K. Hu, Visual pattern recognition by moment invariants, IRE Transactions on Information Theory 8 (1962) 179–187.
[21] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, second ed., Wiley-Interscience, 2001.
[22] M. Kaltenbrunner, R. Bencina, reacTIVision: a computer-vision framework for table-based tangible interaction, in: Proceedings of the 1st International Conference on Tangible and Embedded Interaction, TEI '07, ACM, New York, NY, USA, 2007, pp. 69–74.
[23] TUIO: A Protocol for Table-Top Tangible User Interfaces, Vannes, France, 2005.
[24] UNESCO, Convention for the Safeguarding of the Intangible Cultural Heritage, 2003.
[25] A. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, Content-based image retrieval at the end of the early years, IEEE

Transactions on Pattern Analysis and Machine Intelligence 22 (2000) 1349–1380.

[26] G. Ciocca, I. Gagliardi, R. Schettini, Quicklook2: an integrated multimedia system, Journal of Visual Languages & Computing 12 (2001) 81–103.

[27] G. Pass, R. Zabih, J. Miller, Comparing images using color coherence vectors, in: MULTIMEDIA '96: Proceedings of the Fourth ACM International Conference on Multimedia, 1996, pp. 65–73.

[28] J. Nielsen, T.K. Landauer, A mathematical model of the finding of usability problems, in: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems, CHI '93, ACM, New York, NY, USA, 1993, pp. 206–213.

[29] C. Lewis, J. Rieman, Task-Centred User Interface Design: A Practical Introduction, 1993.

[30] J. Brooke, SUS: a quick and dirty usability scale, in: P.W. Jordan, B. Thomas, B.A. Weerdmeester, I.L. McClelland (Eds.), Usability Evaluation in Industry, Taylor & Francis, London, 1996.

[31] A. Bangor, T. Staff, P. Kortum, J. Miller, Determining what individual SUS scores mean: adding an adjective rating scale, Journal of Usability Studies 4 (2009) 114–123.