Univ Access Inf Soc
DOI 10.1007/s10209-014-0365-0

LONG PAPER

Enhancing education through natural interaction with physical paper

George Margetis · Xenophon Zabulis · Stavroula Ntoa · Panagiotis Koutlemanis · Eleni Papadaki · Margherita Antona · Constantine Stephanidis

Abstract Pervasive computing environments have permeated current research and practice, unobtrusively augmenting existing environments with digital content. The present work, following a pervasive computing approach, proposes a framework to augment an educational environment, being a typical classroom or any studying environment. In this context, the work presented in this paper investigates unobtrusive interaction and support of active educational or studying activities through appropriate context-sensitive information. To this end, passive visual sensing is employed in order to unobtrusively perceive the current context and users' actions, thus providing novel ways to implement natural interaction. The suitability of the proposed interaction technologies and overall approach has been demonstrated through three interactive applications integrated in the framework, each one supporting different interaction techniques and addressing different educational activities. Finally, a user experience evaluation of the three test-bed applications has been carried out, aiming to assess the applicability of the approach and the suitability of each of the proposed technologies to the educational tasks in hand.

G. Margetis · X. Zabulis · S. Ntoa · P. Koutlemanis · E. Papadaki · M. Antona · C. Stephanidis (✉)
Institute of Computer Science, Foundation for Research and Technology – Hellas (FORTH), 70013 Heraklion, Crete, Greece
e-mail: cs@ics.forth.gr

G. Margetis
e-mail: gmarget@ics.forth.gr

X. Zabulis
e-mail: zabulis@ics.forth.gr

S. Ntoa
e-mail: stant@ics.forth.gr

P. Koutlemanis
e-mail: koutle@ics.forth.gr

E. Papadaki
e-mail: elpap@ics.forth.gr

M. Antona
e-mail: antona@ics.forth.gr

C. Stephanidis
Department of Computer Science, University of Crete, Heraklion, Crete, Greece

1 Introduction

The concept of disappearing computers and technologies that "weave themselves into the fabric of everyday life" [1] refers to extending human–computer interaction with conventional objects in the physical world. Although progressive and visionary when proposed, it has now become a common practice reality in many paradigms of human–computer interaction [2]. An important asset of ubiquitous computing environments is their invisibility, allowing users to interact almost at a subconscious level. To this end, ubiquitous computing environments are required to pervade the existing physical environment, carefully augmenting it with digital content, and support users' natural interaction with the environment.

Along the above lines, the work reported in this paper engages the application domain of education, proposing a framework that supports natural interaction with physical artifacts related to education, such as books and pens. With the aim to support several types of educational activities, the proposed framework features three applications for various contexts, ranging from a typical classroom school desk to an augmented study desk and an educational system addressing young learners. The purpose is not to substitute typical education activities, but to augment them

Published online: 26 July 2014

by offering computational support and access to context-related content in the existing environment. In this respect, the framework aims to combine the benefits of conventional education processes employing paper-based learning with those of e-learning and augmented reality [3–5].

In order to address the requirement of invisibility and natural interaction, the framework employs computer vision techniques and, more specifically, passive sensing (vision) and non-instrumented input devices. Computer vision is the preferred method for sensing the environment in ubiquitous and interactive applications, because of its unobtrusive nature and of the wide breadth of information that can be visually retrieved through an ordinary sensor. For instance, the framework supports context identification by (1) perceiving which book has been placed on the desk surface and deploying henceforth the appropriate educational material (e.g., English language, Mathematics, Physics course, etc.) and (2) identifying the currently open book page, thus presenting appropriate information for the current classroom or studying activity (e.g., read aloud a specific text passage, display a video for the current reading activity, provide dictionary information for a word the user is looking for, etc.).

The ability to detect point and contact events on the surface of the desk or the book itself increases the interaction options that can be offered and provides input to the educational applications. Computer vision can be employed to provide touch sensitivity upon virtually any surface unobtrusively, without the use of embedded sensors or markers. In this way, natural interaction is supported by tracking objects in the workspace (e.g., pens, cards, or user fingertips).

By embedding computing in real environments, the physical and the digital worlds, which have been disjoint until now, are brought together. Such fusion enables sensing and control of one world by the other [6]. However, computing infrastructures should not intrude in the environment, either in terms of the physical space or in terms of the activities being performed. The framework discussed in this paper can be integrated in a variety of computing infrastructures, so as to ensure that the appropriate equipment is employed according to each context of use. For instance, framework applications that address typical classroom environments can be embedded in augmented school desks [7], while applications aiming at supporting typical studying activities can be deployed on any desk using ceiling-mounted projectors. Three different audiovisual displays have been employed in this work: a conventional 32″ computer display, a custom wide-screen display integrated in a school desk, and a large projection display.

This work prioritizes vision over speech recognition to unobtrusively obtain user input. The reason is the need to passively acquire information from the environment without overloading the user with uttering commands. In addition, interaction through natural language can be difficult when generalizing to multiple languages or dealing with students of young age. The use of natural language as a means of interaction in education is beyond the scope of this work, though is nevertheless acknowledged as valuable and left for future work.

The remaining of this paper is structured as follows. Section 2 reviews related work in the domains of augmented reality in education, augmentation of printed matter, interaction with printed matter, as well as pervasive and interactive displays. The proposed approach is introduced in Sect. 3, providing an overview of the developed computational framework and architecture and presenting the input sources of visual information. In Sect. 4, the visual techniques adopted for understanding the environment and detecting user actions are described. Section 5 describes three applications that instantiate the proposed approach through different means of interaction and are based on the proposed framework. In Sect. 6, a comparative user experience evaluation of these applications is discussed toward assessing key framework elements and the suitability of different interaction paradigms for different tasks in the educational context. Finally, Sect. 7 summarizes the paper and discusses directions for future work.

2 Related work

This section reviews related work in domains directly relevant to the work reported, namely augmented reality in education, augmentation of printed matter, interaction with printed matter, as well as pervasive and interactive displays.

2.1 Augmented reality in education

The augmentation of educational processes with technology has become a prominent research area during the last decades. For example, Cooperstock's classroom of the future [8] and later approaches [9] envision technology integration into school classrooms, anticipating the potential benefits that may arise for the current educational system. Augmented reality constitutes a basic medium for applying such approaches. Augmented reality learning applications and systems are mainly used for augmenting physical objects with virtual objects and providing proper information related to an educational context of use (e.g., [10–12]). However, one major issue that had to be overcome in order for initial approaches to find application in real educational environments was the need for expensive and inconvenient head-mounted devices (HMDs).

The evolution of technological equipment has driven augmented reality (AR) approaches to a direction toward real environments. Currently, the most prominent AR approaches in education are those based on handheld devices as basic means of interaction, providing information related to real objects or sceneries captured by an integrated camera and presented on the handheld device's display. Some indicative applications following this approach are [13–15]. A more challenging perspective of augmented reality in education includes approaches aiming to provide technological augmentation based only on users' interaction with physical learning assets (books, pens and pencils, etc.). The following sections discuss such approaches raising aspects related to natural interaction, applicability, and educational process and how the proposed framework addresses them.

2.2 Augmentation of printed matter

Since the early 1990s, the idea of digitally augmenting reading and writing, two of the most important everyday life activities, was intriguing enough to trigger the first research efforts in this direction. Wellner's pioneering concept of an interaction continuum between physical documents and their electronic counterparts and the proposed DigitalDesk [16] constitute a prominent milestone of research in this area. Its successor EnhancedDesk [17] built further on the notion of integrating physical and digital documents and aimed at the enhancement of printed information with multimedia in a natural and unobtrusive manner. These works have shown that visually sensing the worlds is an efficient and unobtrusive manner to recognize the documents, which occupy a user's workspace. Still, in EnhancedDesk, the documents have to be physically prepared by bearing a 2D matrix code for recognition. On the other hand, DigitalDesk employs text detection and Optical Character Recognition (OCR) to retrieve text and recognize documents, but also inherits the limitations of OCR, while at the same time it is not suitable for illustrated content, which is typical in educational material. The work reported in this paper uses an alternative approach for printed matter recognition based on a vision process, which tries to match printed papers, placed upon a surface, with their digital images from a closed predefined set. The aforementioned approach provides robust and accurate results even for printed material that contains little or no text.

More sophisticated AR solutions have been subsequently proposed, capitalizing on more recent technologies, such as high-quality 3D graphics. For example, MagicBook [18] offers augmentation of physical books with avatars through VR glasses, implementing a lively storytelling approach. However, such approaches require special editions of educational printed material, which should contain visible markers (e.g., QR codes) in order to recognize pages. On the other hand, the framework proposed in this paper can seamlessly be integrated in any educational environment without requiring modifications of educational assets that are already in use.

2.3 Interacting with printed matter: pointing and touching

Immersive environments provide new avenues in educational research and highlight the value of haptic interaction. Pointing and touching in educational environments, whether natural or augmented, offer a natural means of user interaction. In the context of interacting with printed pages, several interaction devices have been employed.

Digital pens capable of recognizing marks on paper documents facilitated the development of systems supporting annotations on paper. The Anoto system [19] combines a unique pattern printed on each page with a special digital pen to capture strokes made on paper. Similarly, PapierCraft [20] proposed a system recognizing a vocabulary of pen gestures. Furthermore, the Paper++ system [21], besides employing a special electronic pen, requires the use of special paper overprinted with an imperceptible pattern that encodes the page id and location. As in the case of VR, touch-based interaction has been dependent on additional specialized hardware, which often hinders the application of the approach in practice. Proprietary technological artifacts such as light pens, pen with pads, and haptic devices [10, 22] require the installation of sensors or the deployment of particular devices.

Toward alleviating such technical difficulties, some approaches promoted visual sensing but still required the configuration of the environment. Live Book [23] offers natural interaction through the users' fingers, which are recognized using an IR camera and active illumination; still, interaction is limited only to the open page of the augmented book. The "Mixed Reality Book" [24] is another example of augmented book content, page recognition, and interaction via pen, but is also based on markers printed on each page. Intelligent Paper [25] augments foldable prints such as newspapers, maps, and pocket books employing a custom pen device.

All the aforementioned approaches demand special technological means for physical paper augmentation and interaction. On the other hand, the work described here follows an unobtrusive approach for educational environments' augmentation, avoiding the need for technological components that have not been used in these environments till today.

An alternative means of interaction within the context of paper augmentation is through physical cards. Card-based

interaction, which has mainly been explored in the context of educational games, has been claimed to captivate the learners' interest and reinforce motivation [26, 27]. The "Educational tabletop mini games" [28] are played with physical cards, which are visually recognized as part of an Ambient Intelligence classroom [29, 30]. In this work, physical cards are visually recognized, but, also, their spatial arrangement as produced by the learner becomes a means of user interaction.

Recognizing the importance of both physical books and electronic material in the educational process, this paper focuses on physical paper augmentation in the context of current learning practices, such as reading and writing. To provide unobtrusive access to the provided services, specific emphasis is put on natural interaction with the system, adopting touch-based interaction with interactive applications, stylus-based interaction for handwriting tasks, turning book pages and pointing through fingers at interesting page parts (text passages, images, etc.), and card-based interaction for educational games.

2.4 Pervasive and interactive displays

A number of touch-based interactive surfaces exist both as research prototypes [31–36] and as commercial products [37–39]. Most of these systems use a static planar surface as an interaction surface, because it can also be used as a desk.

The first approaches toward augmented interactive surfaces [40, 41] concerned the projection of visual content on convenient surfaces in the environment, such as the walls or the floor of a room. Selecting the surfaces a priori, a projector was steered to display upon the surface of choice. Images were pre-warped according to the orientation of the surface, so as to appear undistorted after projection. Using a more detailed 3D model of the whole scene, projections can appear undistorted at virtually any geometry of surfaces [42]. This way of pervasively augmenting information is still useful, along with conventional displays, as a means of creating pervasive and ergonomic displays for study.

The counterpart of an interactive display is user input. Earlier approaches used electronic devices to provide input, such as the electronic stylus and touch pad [43]. More pervasive approaches such as the aforementioned perceive coarse hand gestures and track an instrumented artifact, a stylus with an infrared beacon, to decrease obtrusion and increase the workspace size [40, 41]. The recent growth of depth cameras enabled better tracking of hand posture and facilitated tangible interaction upon arbitrary surfaces [44]. Still, pointing devices can be particularly convenient in some tasks, particularly if implemented by an ordinary pen that maintains its common use. The evaluation reported in this paper aims at a comparison of such approaches with respect to a range of educational tasks.

Some approaches in the literature support dynamically moving interaction surfaces that can be manipulated by the user. In such cases, the location and pose of the surface itself can avail valuable information to the user interface. In [45], a coarse estimate of the inclination of a handheld surface (a piece of cardboard) provides input to an interactive game. In [46, 47], a similar surface is used to explore maps. In [48], a 2 DOF rotating disk is used as an augmented interactive display whose position can be manipulated. Accurate pose estimation as the user manipulates the interactive surface is essential in all the above cases in order to properly project text and images upon it.

In the desktop context studied in this paper, the augmentation of the study surface is proposed as a means of displaying additional information. Therefore, it is of interest to align the projection with an article of printed matter (i.e., a book or a leaflet), and thereby, the estimation of the book pose is of relevance to how information is displayed. Markers have been utilized in the past to detect the pose of visual documents [18]. The proposed approach is less obtrusive, as it uses the visual content of the document itself to estimate its pose.

3 Enhancing the educational process

The main goal of the proposed educational framework is to enhance the educational process, augmenting physical educational assets (i.e., printed matter) with technological features, in an unobtrusive and user-friendly manner. To this end, the proposed system facilitates education by augmenting basic learning approaches such as reading and writing, independently of the context and the environment that is used for learning. It is of paramount importance that the method integrates seamlessly in the conventional everyday processes of learning, e.g., the classroom resources, and studying material. In this respect, the proposed system can be directly tested and, potentially, adopted in contemporary educational systems without requiring radical changes for its application in school or home study.

In order to achieve this goal, such a system should address the following prerequisites:

(a) It should integrate mainstream equipment that can be easily acquired by educational institutions and students.
(b) It should be easily adopted in the current educational system and support today's and future curricula.
(c) It should be able to provide seamless interaction with any educational asset (e.g., books, pencils, pens)

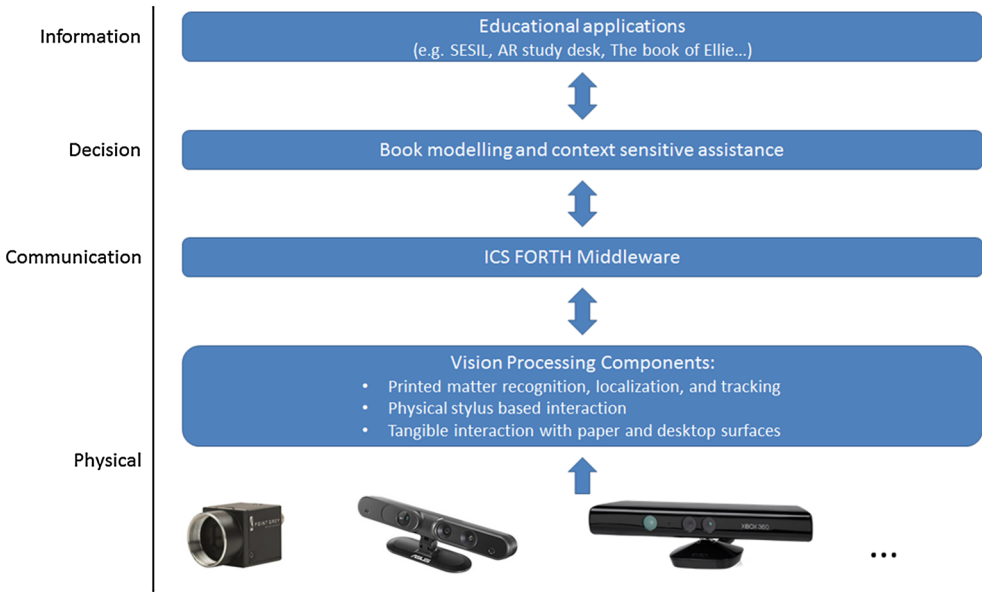


Fig. 1 Framework architecture

adopting currently used learning practices (reading, writing, touching), avoiding the enforcement of extraneous technological devices such as computer mouse, haptic devices, or proprietary stylus.

The proposed system has been designed as an Ambient Intelligence system that can be used for transforming any conventional learning environment into a smart environment. According to Cook et al. [49], any smart environment can be adequately decomposed in four fundamental layers: physical, communication, information, and decision. Each layer performs a different role in the environment, facilitating diverse operations and addressing specific requirements.

As illustrated in Fig. 1, the system's overall architecture comprises the four fundamental logical components of an Ambient Intelligence system. The physical component consists of vision processes that receive video feeds of various types (high/low resolution, conventional or depth images). Vision components recognize and track physical objects (e.g., book pages, pens, hands, or fingertips) in order to provide interaction data. The complexity of using the vision-based systems (or sensors-based systems in general) is counterbalanced by the adoption of the "ICS FORTH Middleware,"[1] which provides the necessary functionality for the intercommunication and interoperability of heterogeneous services hosted in the environment. Furthermore, a decision component has been developed providing educational book modeling and facilitating the necessary context-sensitive information provision to educational applications that integrate the proposed framework.

[1] FAMINE (FORTH's AMI Network Environment), has been implemented in the context of the ICS—FORTH's internal RTD Programme 'Ambient Intelligence and Smart Environments'.

Computer vision is a passive method that requires minimal intervention to observe the environment and, thus, facilitates the seamless integration of such approaches in the classroom or at home. The case addressed in this work is the interaction of a user with physical documents, whether a book or an illustrated card, on a desktop environment. Actors in this environment are the hands of the user and, potentially, a pointing device. The particular card or page that is currently open provides context information about what is the current topic of study or its conclusion, when the book is placed aside. In Fig. 2, the information flow produced by the computer vision processes is shown. These processes detect and recognize meaningful events occurring in the environment and track the 3D locations of the objects that participate in them. They can be classified according to the following two categories of provided information

- Recognized pieces of content such as which book page is open or which card was presented provide contextual information regarding the exact book page currently studied. The recognition needs not to be rapid but needs to be robust, because the page may be occluded by hands.
- Accurate estimation of the pose of a pointing device, or the finger itself, can be used to indicate a particular figure, sentence, or word in that page. Real-time estimation enables the use of natural pointing gestures in system interaction. This more time-dependent information is provided by tracking modules.

Having a modular way to provide this information allows the dynamic composition of a system's "perceptual abilities," thus allowing to investigate the efficacy of

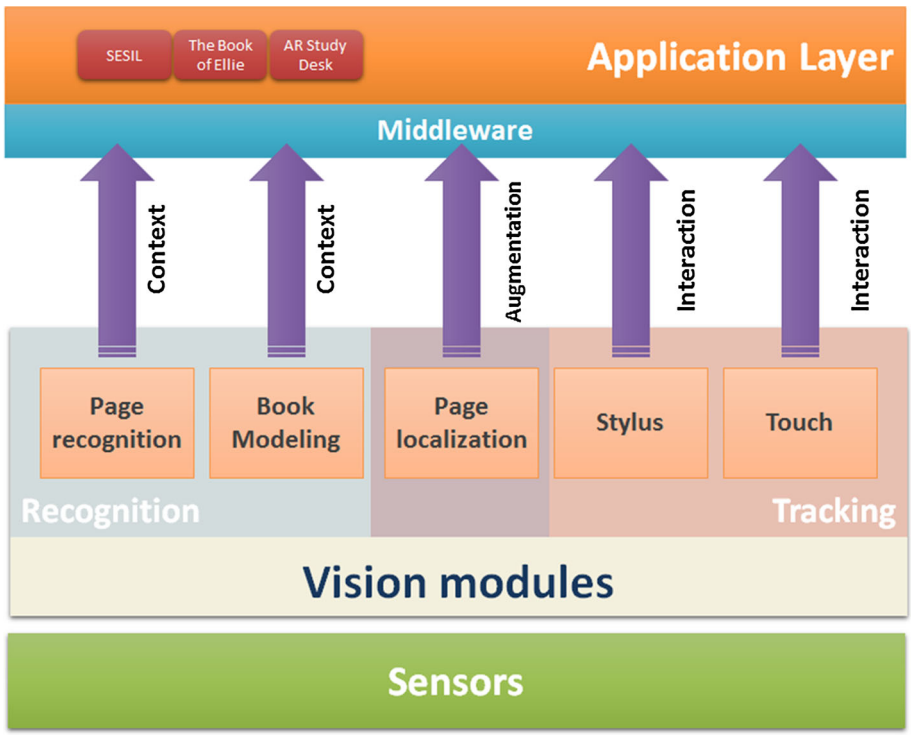


Fig. 2 Information flow for computer vision processes

alternative approaches, such as using a pencil or fingertip as a pointing device. Toward this end, three systems that investigate alternative interaction paradigms were designed, implemented, and evaluated, offering valuable insight regarding the technical difficulties, but also regarding the acceptance (intuitiveness) and suitability of different interaction technologies and presentation approaches, i.e., suitability of projecting versus showing on screen for each application. To this end, with the goal of assessing the usability and unobtrusiveness of the three systems in relation to the interaction techniques employed and their suitability for the educational process, a usability evaluation was planned and conducted.

4 Understanding the environment

The perceptual modules of a system aim to find information in the environment that are useful to the educational application. In the current context, the recognition of pages (or illustrated cards, as a robust contextual cue) has been selected as most relevant. Besides recognition, the locations of documents on a desk can be used to augment the environment and find which piece of a page was pointed at or underlined, either using a fingertip or using a pen. Finally, the recognition of the book content, through book modeling and context-sensitive assistance, provides the contextual information to the educational applications.

4.1 Printed matter recognition, localization, and tracking

A key vision functionality concerns the detection of the presence of a book in the workspace, its location, as well as the recognition of the currently open page. The approach adopted assumes that the book is already known to the system and its electronic representation is provided.[2] Geometric information regards the 2D pose of the book upon the desk, that is, its 2D location and 2D orientation. Context information concerns the id of the two book pages that are typically open.

4.1.1 Page recognition

In order to recognize book pages that are open on the physical desktop, the module accesses the electronic representation of the book. In a nutshell, the module searches the electronic book pages and compares them with the live image acquired from the camera of the system, to find out whether any such page appears in this image. In the electronic representation, each page has been previously imaged in a planar arrangement (either scanned or synthetically generated, i.e., as a PDF document). A feature-based representation is maintained for each page, obtained by detecting SIFT keypoint features [50] in the image of

[2] For school books this is a reasonable requirement, as most are available electronically.

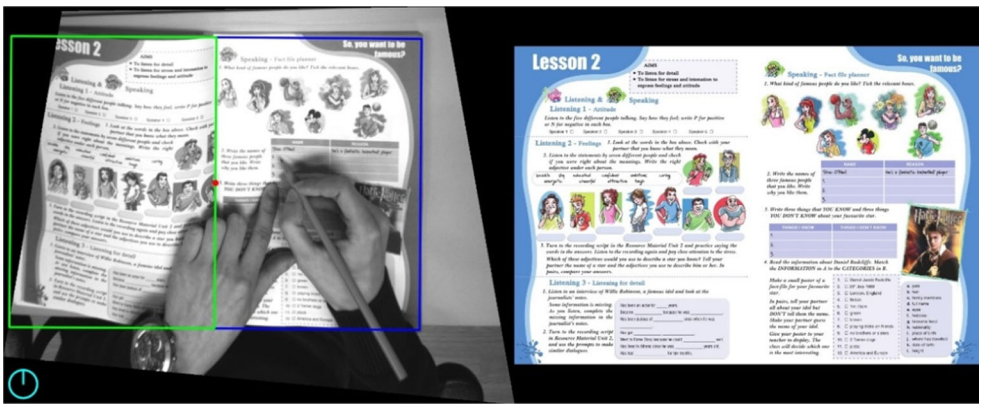


Fig. 3 Recognition and localization of book pages. On the *left*, an image acquired from the system camera is shown. On the *right*, the matched pages in the electronic representation are shown. The *left* image is annotated with the estimated spatial extent of each page (*green*, *blue*) and the estimated book orientation (*cyan*). The outcome maps physical points of the surface of the particular page to coordinates in its electronic representation

each page. The detected keypoints are associated with the electronic page where they were detected. In other words, only the features of each page are maintained in the electronic representation.

At run-time, the SIFT keypoint features are detected in the live image acquired from the system camera. These features are matched with the ones in the electronic representation of the book, on a per page basis; matching features are found using the conventional method described in [50]. For each candidate page, a "hit ratio" is computed as the fraction of the number of the matched features over the number of the total page features. The page that provides the highest hit ratio is considered as the matching page. To optimize the search, consecutive and preceding pages to the current one are considered first, and if they provide a very high ratio, they are immediately considered as matching. As keypoints are local features, page recognition is robust to occlusions because the entire set of page features is not required to be matched to achieve recognition (see Fig. 3). The proposed recognition approach can recognize pages that undergo (approximately) as much as 50 % occlusion [51]. Still, if a page is more severely or even entirely occluded, it can be implicitly recognized if its side page is successfully recognized. The recognition approach is presented in further details in [51].

4.1.2 Page localization

The estimation of book location and orientation utilizes the synthetically created helper image W. When an image is acquired from the system, it is initially warped to the plane of the desk, providing image W. This is achieved by estimating the pose of the camera with respect to the plane of the physical desktop through conventional camera calibration, during system installation. Thereby, image W is an orthocanonical representation of the current desktop appearance, where distances are in direct analogy to the physical distances on the plane of the desktop. As W is orthocanonical with respect to the coordinate frame of the desk, the estimations of book location and orientation can be directly mapped to the physical spatial desktop.

The spatial arrangement of keypoints provides a basis for the localization of the book on the desk. The matches obtained for the current page establish point correspondences between the electronic page and the page as displayed in W. Using these correspondences and approximating the 3D shape of a page with a planar patch, the homography that maps an imaged page to the electronic one is estimated through a RANSAC procedure. This procedure also outputs the subset of confidently correct correspondences, which are called inliers. Using the estimated 3×3 homography matrix and the (known) book dimensions, the corners of each scanned page are predicted in the 3D world and the acquired image. The locations of these four predicted corners mark the spatial extent of the book pages in W. Procrustes analysis is then employed [52] upon the confident (inlying) correspondences to estimate the 2D orientation of the book in W and, equivalently, on the desktop too. The process is performed for each page that is visible on the desktop (typically 2), as each page may have a different orientation in the 3D world (see Fig. 3). Approximating each book page with a plane was found to be sufficient. This is due to the robustness of the RANSAC procedure which, in essence, returns the plane that best fits the planar portion of the page, which is also the part of the page available to the user for interaction.

The module is implemented through parallel programming and executed on the GPU of the computer. The frequency of operation is approximately 7 Hz on a conventional computer and GPU. At each frame that a book

is found, the module outputs an event with the ids of the two (or one, i.e., if the cover is matched) recognized pages along with the coordinates of each page in the desk's coordinate system. The coordinates are henceforth utilized to support tangible interaction with the book. An error is produced in the rare case that two non-consecutive pages are recognized (typically, this occurs when illumination is poor). Finally, locations and orientations are tracked through Kalman filtering [53] to cope with transient recognition failures.

4.2 Physical stylus-based interaction

A module visually estimates the 3D poses and the endpoints of a thin cylindrical physical object, such as a stylus, which is manipulated by the user [see Fig. 4 (left)]. The stylus may be of unknown size and may be occluded totally or partially. Most importantly, the stylus needs not to bear any electronic device or even markers, but only be of a known color. The response of the method is real time independently of the used hardware.

The method utilizes multiple synchronous calibrated images of the object to increase accuracy and deal with occlusions. In each view, it is segmented and modeled as a line segment that approximates the projection of its major axis on the image plane. When segmentation is successful in 2 views, the object's size and pose can be estimated. If more views are available, they are combined to increase accuracy.

In Fig. 4 (right), the case of two views imaging the stylus is illustrated. Each camera i is located at K_i and has a projection matrix P_i. The image from each camera is compensated for lens distortion directly after its acquisition, forming image I_i. The output is the 3D line segment, represented by its endpoints $e_{1,2}$. The method has the following steps:

1. Segmentation. Each image I_i is binarized into image M_i to extract the wand. Segmentation may contain errors, such as spurious blobs, while the wand may not be fully visible due to occlusions. Using a wand of characteristic color, selected to be scarcely encountered in the scene, M_i is obtained as follows. A color similarity metric [54], robust to variations of illumination conditions, yields a similarity score per pixel. Each result is thresholded to generate binary image M_i (Fig. 5, middle).
2. 2D modeling. The wand is sought in each M_i, using the Hough transform (HT) [55], which yields 2D line l_i. Due to perspective projection, the stylus does not exhibit constant thickness in I_i. A thinning process estimates the width of the stylus in the image and, based on its value, finds the medial axis of the wand, which, in turn, is provided as input to the HT. Finally, a line segment grouping process upon l_i determines the endpoints of the wand in I_i (Fig. 5, right).
3. 3D pose estimation. The 3D line L where the segment lies is estimated as the 3D line that minimizes its re-

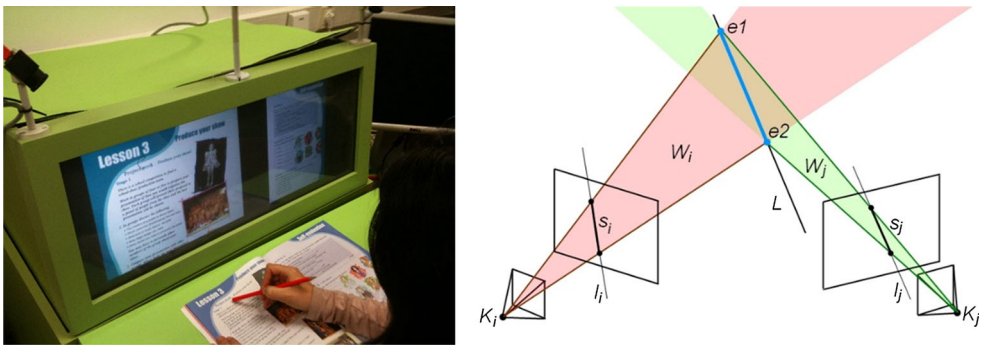


Fig. 4 *Left* Using a stylus to point on locations, circle words, and underline sentences, on the surface of a book. The system employs three cameras that overlook the user workspace. *Right* Illustration of the method's geometry for estimating the location and orientation of the stylus from a pair of cameras (see text)

Fig. 5 Segmentation and 2D modeling of the stylus in a color image (see text). Original image I_i (*left*), M_i (*middle*), and M_i with l_i and s_i superimposed (*right*); the *red line* is l_i and the *blue* segment is s_i

projection error to the detected line segments in the acquired images; the SVD-based minimization is described in [56]. Endpoint estimates are obtained by triangulation of their 2D observations. In this process, outlier elimination is crucial as, due to occlusions and segmentation errors, the object may not be fully visible in all views.

4. Motion estimation and tracking of the stylus improve the accuracy of pose estimation and increase robustness by correcting transient errors, i.e., when the stylus is lost or pose estimation is inaccurate for a few frames. The trajectory of the wand is tracked over time by conventional application of a Kalman filter.

The method supports real-time interaction by massively parallelizing the required computation and implementing most of it in the GPU of the computer and, in a conventional setup, provides less than 1 cm of error in the estimation of pointing or contact location. More details on the implementation of the method and its evaluation can be found in [56].

4.3 Tangible interaction with paper and desktop surfaces

This vision module detects and localizes the physical points on a desktop surface or a book that are in contact with the fingertips of the user. Two approaches are proposed, the first concerning fingertip touch detection and localization on the planar surface of the desktop, and the second concerning the curved surface of a book.

4.3.1 Desktop surface

Touch detection on the desktop surface is performed as follows. As mentioned in Sect. 4.1, the pose of the camera relative to the desktop surface is a priori estimated through calibration. Let E be the equation that represents this plane and D the depth image acquired by the RGBD sensor. The pixels in D are transformed into 3D points. For each of these points, its distance to plane E is computed. As in [57], two thresholds are used to detect contact. The first (*dmax*) indicates if a pixel is closer to the camera than the disk surface. However, only this constraint will include points belonging to the user's arm as well. The second threshold (*dmin*) eliminates points that are overly far from the surface to be considered part of object in contact.

The aforementioned double thresholding supports gathering 3D points within a specific distance range from the desktop surface. To accurately estimate touch events, *dmax* should be as small as possible. On the other hand, due to sensor noise, when the value of *dmax* is very small (i.e., $\approx$1 cm), spurious touch points are detected. This noise is tackled in a two stage process. At a first stage, *dmax* is set at a desired precision ($\approx$0.5 cm) and an appropriate value on *dmin* (i.e., 10 cm) is used to ensure that the detected points correspond to the hand of the user. Then, a 3D connected component labeling on these points is performed, keeping only the largest components under the assumption that they correspond to user's hand(s) and not noise. Points grouped to small components are considered noise and removed. At a second stage, the threshold is doubled again, but using a stricter value on *dmin* (i.e., 3 cm). The remaining points belong to parts of the user's fingers near fingertips. The 3D connected component labeling is performed once more to group these points into connected components (see Fig. 6).

Besides the fingertip area which is in contact with the desktop, a single point of contact is also estimated. This point is estimated as the projection of the centroid of the connected component on the desktop plane. To facilitate application development, the module transforms real-world coordinates to those of a "computer display." In accordance, contact events are re-presented similarly to the indication events of a pointing device (i.e., 'click' for the mouse).

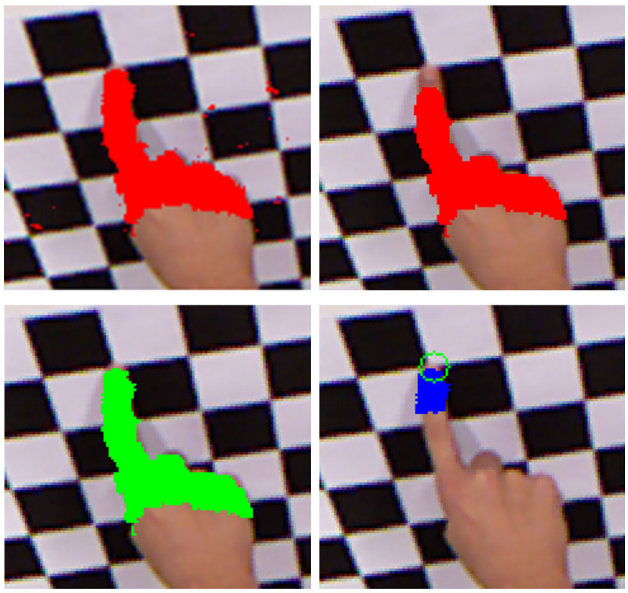


Fig. 6 Detecting 3D points of a user hand above a planar surface. *Left* to *right* **a** the first stage yields points within the range of *dmax* = 0.5 cm and *dmin* = 10 cm, projected on the corresponding sensor's image with *red color*. Spurious points are present due to the small value of the *dmax* threshold. **b** If the value of *dmax* (i.e., to 1 cm) raises, the detection of critical points near the fingertip would fail. **c** To suppress noise but also retain these critical points, the largest connected component in the result of the first stage is found. **d** In the second stage, points belonging to the largest component are thresholded again to a finer range (*d*max = 0.5 cm, *d*min = 3 cm) now providing points at the fingertip; the *green circle* shows the point of contact

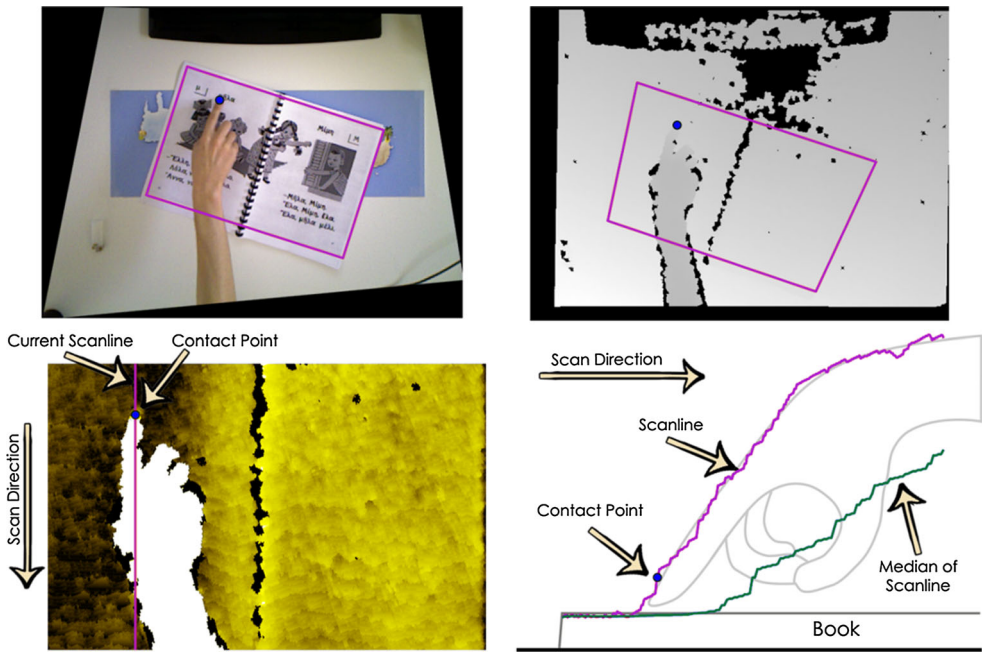


Fig. 7 Contact detection on the surface of a book. *Top left* estimation of the book boundaries in camera image. *Top right* The portion of the depth map used to generate the height map. *Bottom left* the generated height map, displaying a top view of the book. *Dark* pixels are closer to the workspace surface than lighter ones. *Purple line* indicates the scanline used to generate the *bottom right* image. *Bottom right* schematic overview of the contact detection method, from a side view. The *purple curve* corresponds to the purple scanline from the *bottom-left* image and represents distance from the workspace surface. The *green curve* corresponds to the median of the scanline. As the scan progresses from *left* to *right*, more samples from the scanline are included in the median computation

4.3.2 Book page surface

A different approach is followed to perform touch detection on the book surface, as it is typically curved. Using the estimates of book location and orientation, the boundaries of the book are predicted. The corresponding portion of the depth map is then extracted and transformed to an orthocanonical view with respect to the book's coordinate frame. The book locations estimated in camera coordinates are transformed into the desk's coordinate frame. Using this information, a height map of the book's surface is generated. Each pixel of this map shows the distance of the corresponding book surface location from the workspace surface. When the user points at a location on the book, the hand creates an increment in the height map. The contact point is detected by finding the location where this increment starts to occur, that is, where a discontinuity in the, otherwise smooth, book surface is detected. This location is found by scanning the columns of the height map iteratively, looking for the topmost height discontinuity. In iteration i, the median height for the topmost k_i pixels of each scanline is computed. For each scanline, the height at the k_i-th pixel from the top is compared with the median value for this scanline. If a deviation larger than threshold q_L (10 mm) from the median is observed, a contact point is assumed to be found. In order, though, not to spuriously detect as in contact a finger that hovers above the page, this deviation is required to be less than q_H (17 mm). If no contact is found, k_i is increased by 1 and the method proceeds to the next iteration. The contact point is tracked through Kalman filtering to reduce jitter and recover from tracking failures. The process is illustrated in Fig. 7.

4.4 Book modeling for context-sensitive assistance

The applications presented in this paper use a recognition library that contains all the pages of the books used in the evaluation described in Sect. 6 in digital form (Portable Digital Format—PDF). The recognition library has been enriched with educational meta-data for each stored page. In order to decide the information that the system will provide to an application after a user's specific interaction with the environment (touch a page area, write something, etc.), every page of the printed matter used is classified according to the educational asset it provides (e.g., reading text, image, comprehension exercise, etc.), defining in this way the interaction and content assistance needs of the user. For example, when a user tries to accomplish a multiple choice exercise, each application is fed, in real time by the system, with hints and explanations about the specific section of the exercise which is in the user's current attention area.

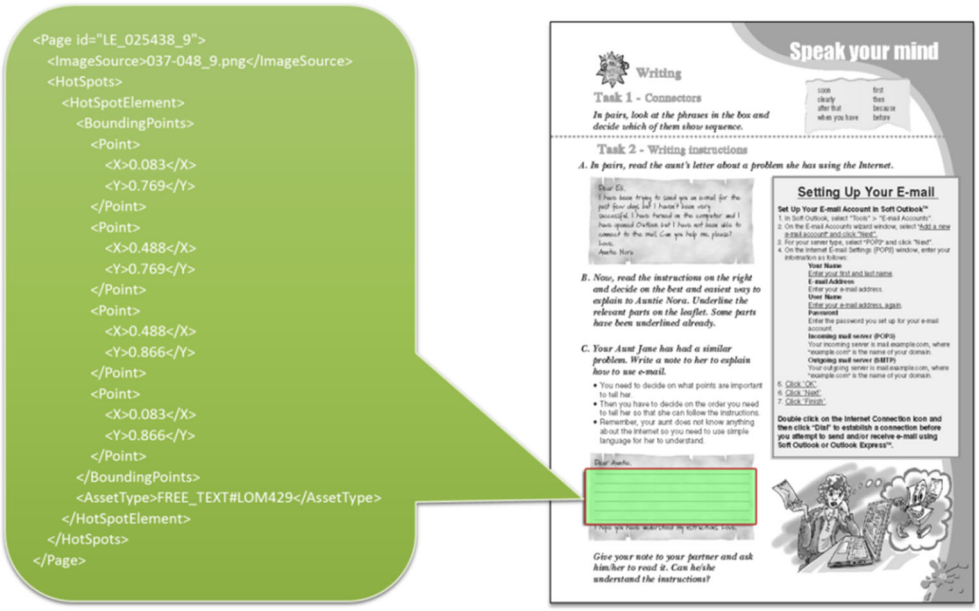


Fig. 8 Book page meta-data information

The classification of printed matter is contained in an XML description stored in a recognition library, in which the interactive area (hot spots) and their types are defined. Figure 8 (left) depicts the meta-data file of a school book page (Fig. 8 right) that contains a free text field of a fill-the-gap exercise. As it can be observed, every page in the recognition library is referenced by a unique id and is accompanied by its digital image path. This image can be displayed on any interactive screen near the physical paper or directly on it, using a video projector, enabling the user to interact with the hot spots of the digital form of the page. Furthermore, every page can contain a number of interactive hot spots denoting its interesting areas. Every hot spot of a page is declared by four coordination points (normalized in order to be independent of the page size), representing the four corners of the hot spot's bounding rectangle and the educational asset type that this hot spot denotes.

Each asset type is assigned to a number of accepted gestures (e.g., circle, underline, etc.) and/or writing text. In particular, the assigned writing text to a hot spot can be either a closed set of words or any word that is contained in a vocabulary that may be used (e.g., Open Office Dictionaries[3]), according to the asset type.

Anytime a page hot spot is engaged by the user's handwriting, the content-sensitive assistance module evaluates that input and, according to the type of educational asset that is represented, selects the appropriate information related to this asset for the specific users' input and propagates it to the corresponding educational application.

[3] http://wiki.services.openoffice.org/wiki/Dictionaries.

In order to provide context-sensitive information, the book model component interoperates with ClassMATE [9]. Through the ClassMATE framework, functionalities such as user-profiling, content-classification, content-discovery, and filtering are available.

5 Integrated systems

This section discusses three example AR systems that have been implemented in order to demonstrate the proposed framework and provide real-world test beds for evaluation.

5.1 SESIL

SESIL [51] is an educational system that incorporates the proposed framework in order to provide stylus-based interaction in different spatial arrangements, such as large interactive surfaces featuring a display with multi-touch capabilities (i.e., for use in a library or at an exposition) provided that the cameras, needed for pages and stylus recognition and tracking, are positioned appropriately.

SESIL aims at enhancing reading and writing activities on physical books through unobtrusive monitoring of users' gestures and handwriting and the display of information related to the current users' focus of attention. Additionally, SESIL exploits educational meta-data on the book's content to decide at run-time the type of additional information and support to be provided in a context-dependent fashion.

In more detail, the SESIL system consists of a desk that is being overlooked by a set of three high-resolution cameras placed above it. A nearby large display runs an

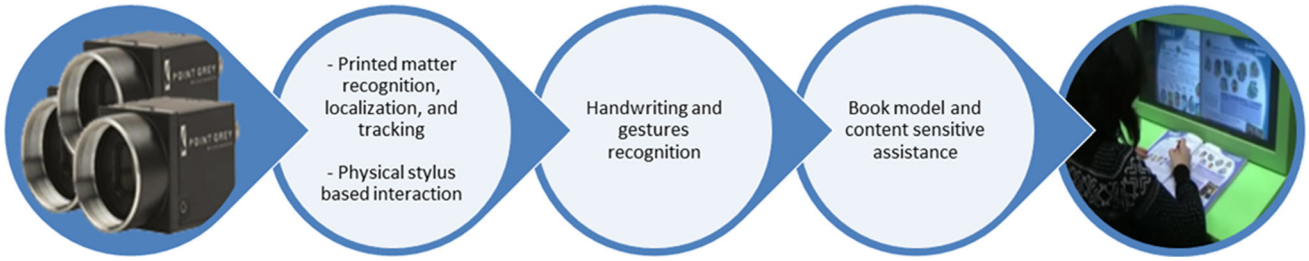


Fig. 9 SESIL components chain

educational application that provides content-sensitive information to the users, based on their stylus-based interaction with a school book.

Figure 9 depicts the components chain of SESIL, designed to facilitate information flow in real time, so that system interactivity is served. In more detail, the images acquired by one of the three cameras are processed for localizing the book on the table and recognizing the particular page that the book is open at. In parallel, the other two cameras are used for the 3D orientation and location estimation of a stylus, while contact information between the stylus and the book or the desk is also captured. When a user interacts with a specific area of the open pages of the book, SESIL provides context-sensitive assistance to the application running on the nearby display.

Stylus input is additionally used to recognize gestures and handwriting. This module takes as input the pose of the stylus and identifies whether the user is performing a gesture or is writing text. The recognizable gestures are strongly correlated with the natural reading process (e.g., underline text, circle a word, etc.), ensuring that standard studying habits and practices are supported. The recognized words or gestures, along with their position on the book's page and their size, are provided to the book modeling and context-sensitive assistance component.

The gestures that are recognized and used by SESIL are:

- Basic geometry (e.g., draw specific geometrical shapes)
- Multiple choice (e.g., square, circle, or tick the correct answer among a number of available choices)
- Fill-in the gap with multiple choice answers (e.g., circle the correct answer to be filled-in a sentence)
- Word search (e.g., identify words in a matrix of letters, given a list of their definitions, by striking through the word's letters)

In addition, SESIL recognizes and isolates gestures, throughout the students' handwriting process, in order to be able to identify specific triggers, such as the deletion of a word that a student has written (identifying the Scratch Out gesture) or the accomplishment of a sentence (identifying the Tap gesture), which means that a student has finished providing input and the system can therefore further process it.

In the context of the evaluation conducted, SESIL was installed and integrated in the augmented school desk described in [7], taking advantage of the desk's existing infrastructure and augmenting it with the additional cameras required by SESIL in order to recognize pen gestures and handwriting.

5.2 AR study desk

The AR studying desk is a newer version of the interactive desk described in [58], aiming to augment physical books with digital information. Figure 10 illustrates the AR study desk's components chain.

The AR study desk consists of a standard definition projector and an ASUS Xtion Pro, both overlooking the surface of a desk. The images acquired by the color camera of the Xtion are used for printed matter recognition and its localization on the desk surface, while the images acquired by the Xtion's depth camera are used for detecting users' finger touch on the printed matter or the desk.

The AR studying desk provides context-aware multimedia and interactive applications related to the content of the open book page. Such content is dynamically displayed to enrich the contents of the book page currently open and is aligned, in real time, with its 2D orientation upon the desk.

Technically, augmentation is supported by the projector-camera calibration. Given the coordinates of the book or the stylus in the desk coordinate frame, this calibration is used to predict the coordinates of the projector pixel that will illuminate the corresponding region or point of interest.

The content-sensitive digital content provided can be classified as follows (asset types):

- Images, optionally followed by informative text
- Videos that have been stored in the system
- Images and videos from online Web sites (e.g., Google[4] and YouTube[5]) that are being collected at run-time according to the user's interaction

[4] http://images.google.com.

[5] http://www.youtube.com.

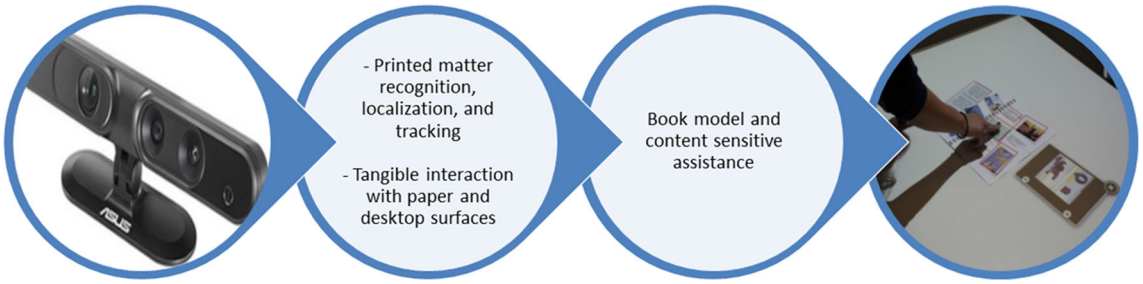


Fig. 10 AR study desk components chain

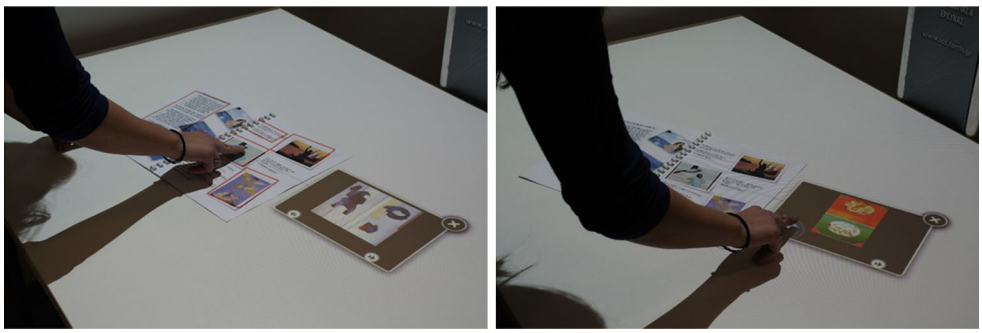


Fig. 11 *Left* the user clicks on a hot spot and a slideshow application opens laterally to the open page playing related videos acquired at real time from a public image Web site. *Right* the user rotates the book, and the content rendered is updated so that the video keeps its alignment with the book, following its motion

In more detail, every printed page that is included in the system's library has been stored in digital form (PDF) and annotated with "hot-spot areas" that play a role in user–page interaction. This annotation takes place using the book modeling and content-sensitive assistance component. When a page hot spot is engaged by the user's interaction, the system evaluates that input and, according to the type of the asset that is represented, selects the appropriate supportive applications and displays them on or near the active book's page. For example, in Fig. 11 (left), the user has clicked on a hot spot and an image slideshow related to it is rendered juxtaposed to the open page. If the user rotates and moves the book on the desk, the rendered application follows it maintaining its alignment with the book (Fig. 11 right). However, users can move applications rendered on the desk to a more convenient location by dragging them with their finger.

The rendered images can be controlled using a small set of gestures. For example, if a user wants to see the next image, he/she can slide his/her finger from left to right on the table, while if he/she wants to see the previous image, he/she would slide his/her finger to the opposite direction.

Furthermore, the system presented provides a note-taking facility, by writing on a free area of the table using a soft keyboard displayed next to it. The user is also able to associate notes with parts of the open page by circling them. Furthermore, he/she is able to email or post the notes to his/her Facebook and Twitter account.

5.3 The book of Ellie

The "Book of Ellie" [59] is the augmented version of a classic schoolbook for teaching the Greek alphabet to primary school children. The book introduces alphabet letters and their combinations by increasing the difficulty level. For each letter or letter combination, relevant images and text involving the specific letter(s) are provided. The short stories for each letter are structured around dialogues and activities of a typical Greek family, with the protagonist being Ellie, one of the four children. In the augmented version of the book, Ellie has become an animated character, constantly available to assist the young learner by reading phrases from the book, asking questions or providing advice.

In terms of setup, the system consists of a television screen (32″) for visual and audio output, an "Asus Xtion Pro" RGBD camera, and a PC running the software. The RGBD camera is used to recognize and localize book pages and cards, as well as detect and localize fingertip contacts on the book and table. The physical book and paper cards (e.g., depicting letters, simple objects, or animals) are interactive components of the system. The system's main components are depicted in Fig. 12.

Furthermore, the system integrates the printed matter recognition and tracking component in order to detect the presence of known books or printed cards on the desktop, recognizes them, and estimates their location and orientation on the desktop coordinate frame. Recognition is based

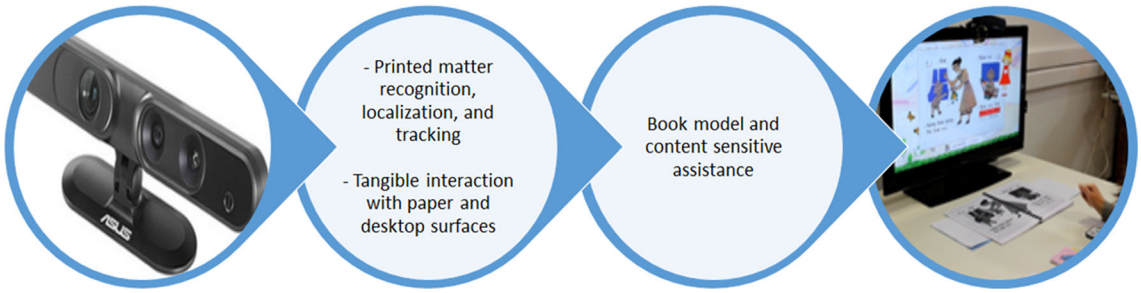


Fig. 12 Components chain of the book of Ellie

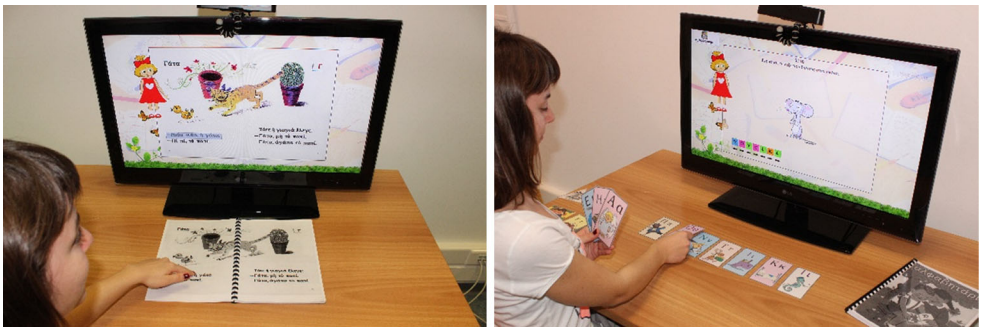


Fig. 13 *Left* book mode interaction: the user has just pointed at a phrase in the physical book, which is highlighted in the screen as Ellie reads it aloud. *Right* spell-the-word card game

on the book model and context-sensitive assistance module. Additionally, tangible interaction on physical surfaces component is used in order to detect users' finger contact with book regions.

The application supports two learning modes, (a) reading mode and (b) game mode. When the reading mode is enabled, the student can turn the pages of the book as she likes. The system monitors the child's actions and displays the electronic version of the open page to a nearby display. If the child touches a sentence of the physical page, then the virtual character Ellie starts reading this sentence aloud, while it is highlighted on the electronic page that is presented on the display (see Fig. 13 left).

Swapping to Game mode (see Fig. 13 right), the child is introduced to an educational card game, which acts as a recapitulation of the letters that have been taught and a teaching tool for the spelling of some basic words. During the game, the child is asked twenty-four questions randomly chosen from two categories, asking the child to: (1) select a card with a picture that corresponds to a word which begins with a specific given letter (e.g., Lion for "l") and (2) spell the word represented by a given picture, using cards representing letters (e.g., Lion is spelled by the cards representing the following letters "l," "i," "o," "n").

Mode swapping is achieved by placing special purpose (utility) cards on the desk, one for each mode. The child can activate a mode at any time and resume its interaction from where it was left (i.e., resume reading from the last page that was open or continue answering the last question that was not successfully completed).

The vision module detects and recognizes the cards that appear on the table at any given time. In addition, it characterizes the spatial arrangement by which the cards are laid out on the table, in order to detect whether they are in a linear spatial arrangement and properly oriented. The orientation of cards is required to be compatible with that of the line. The cards are considered to be properly oriented if their orientation is not more than 20 degrees different than that of the line. In this way, arrangements that include misaligned cards (e.g., upside-down) are not recognized as valid.

6 User experience evaluation

The main evaluation goal was to identify whether the proposed framework is unobtrusive in the context of the educational procedure. Additional evaluation goals included the comparison of the available interaction techniques. Usability evaluations had previously been carried out for all three systems [51, 58, 59].

To this end, a within-subjects [60] evaluation experiment with sixteen participants was planned and carried out, during which each participant used and evaluated all three

Table 1 Evaluation participants' characteristics

	Overall (%)	Students (%)	Parents (%)
Gender			
Male	50	41.5	75
Female	50	58.5	25
Educational S/W expertise			
Low	12.5	16.5	0
Medium	43.75	33.5	75
High	43.75	50	25

systems. The within-subjects test was preferred since it is considered as more appropriate when the evaluation aims to compare different variables, as the various interaction techniques used in the three systems. In order to address any bias in the results due to learning effects, the order in which participants evaluated the three systems was varied.

6.1 Participants

Participants were selected to be familiar with computers, nevertheless their expertise in the use of educational software and multimedia varied. Furthermore, twelve of the participants (75 %) were Computer Science students, while the remaining four participants (25 %) were recruited as parents of young children who use educational software for their children. It should be noted that the majority of participants were selected to have some experience with educational software and multimedia, since the main evaluation goal was to assess the unobtrusiveness of the system as perceived by the users. Therefore, a comparison of the participants' experience with the proposed system versus standard educational and e-learning software would provide a reliable indicator. Information about the participants' profiles is presented in Table 1.

6.2 Procedure

An important parameter affecting the evaluation methods that were employed during the experiment was the research question that should be answered in this evaluation. The main evaluation goal was not to identify specific usability issues of the three applications; it was instead to understand how users experience the overall educational procedure with the support of the AR systems, as well as to find out users' satisfaction regarding the supported interaction techniques. Since collecting quantitative data was not a primary concern, the evaluation was carried out by observation [61] of participants' interaction with the systems, taking notes of the errors that were observed and the difficulties met. During the process, the participants were encouraged to express their thoughts aloud [62], thus

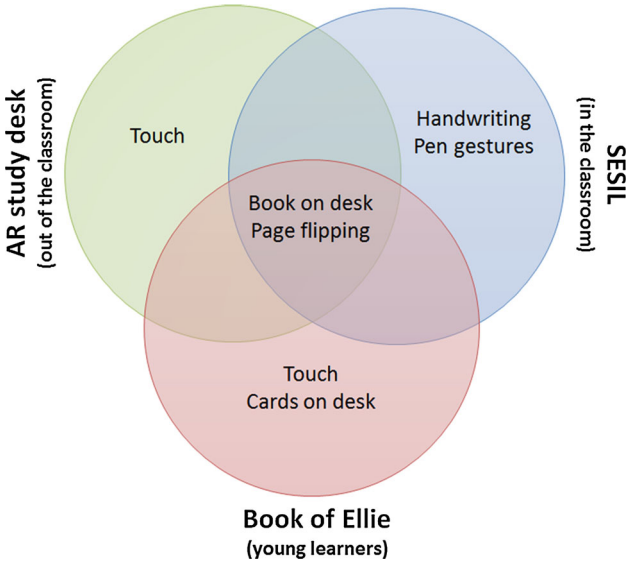


Fig. 14 Interaction modalities and educational contexts for each system of the proposed framework

assisting the observer's note-taking process. After using each system, participants were asked to fill-in a short user experience questionnaire, while the evaluation concluded with a semi-structured interview focusing on the participants' view of the systems. The interview aimed at clarifying any interaction issues or usability problems that occurred during the test, and acquire further insight into the users' opinion about the three systems.

During the test, the participants were welcomed and introduced to the systems, their content, and the various interaction modalities supported. An important concern which was raised during the evaluation planning and preparation phases was the heterogeneity of the three systems in terms of content and interaction techniques. In more detail, both the content and the interaction modalities for each system were selected, so that they would best fit different educational contexts, as displayed in Fig. 14. SESIL was designed having in mind a typical classroom activity of reading and exercise-solving. The AR study desk targets more exploratory activities where the learner receives information about a topic alone or in collaboration. Finally, the Book of Ellie addresses younger children and edutainment activities, where learning is achieved by play. Design decisions were not revealed to the evaluation participants, in order to elicit their opinion regarding the association of interaction techniques with specific educational contexts and avoid bias.

Furthermore, aiming at obtaining qualitative data, users were not provided with detailed scenarios structured around specific tasks. On the other hand, a free exploration of the system might lead to long evaluation sessions and unsuccessful interviews with exhausted participants.

Therefore, an indicative scenario highlighting the main system elements that the user should explore was created and used by the evaluation observer to guide the experiment. Scenario overviews are provided in Table 2.

The user experience evaluation questionnaire comprised two sections: a section with statements regarding users' experience with the system and a section with pairs of contrasting characterizations for the system. In more detail, participants had to specify their agreement on a scale from 1 (strongly disagree) to 5 (strongly agree) with each one of the following seven statements:

1. "The system was easy to use"
2. "I didn't need a lot of practice until I learned how to use the system"
3. "The system was awkward to use"
4. "The system disrupted my workflow and disoriented me from my learning goals"
5. "Interaction with the pen/by touch/with cards was easy to achieve"
6. "Such a system can make studying more effective"
7. "Such a system can make studying more pleasant/enjoyable"

In the case of systems employing more than one interaction technique, the relevant question was asked once for each interaction technique.

Furthermore, the system characterizations (Pleasant—Unpleasant, Interesting—Boring, Straightforward—Cumbersome, Predictable—Unpredictable) ranged on a scale from 1 to 7, with 1 representing the most positive attitude (e.g., pleasant) and 7 the most negative (e.g., unpleasant).

Table 2 Scenario overviews

System	Scenario description
SESIL	The user was asked to open the book at the first page of a chapter (thematic unit). There he/she was instructed to read the content, browse the next couple of pages, and find additional information for specific words that might be of interest. Then, the user was asked to move to the chapter's exercises and solve a fill-the-gap exercise with multiple choice answers, as well as to answer an exercise by providing handwritten input
AR study desk	The user was instructed to select one of the available leaflet books and initiate interaction with the system. Then, he/she was asked to browse through the pages of the book and select three specific points of interest (text passages or images) for which he/she would like to receive additional information
The book of Ellie	The user was asked to browse through the book pages and select text passages or letters of the alphabet of interest. After a short period of interaction with the book, the user was urged to switch to game mode

The participants were asked to provide a rating indicating how they felt about the system they had just interacted with.

Finally, interviews aimed at further investigating the participants' view of the interaction and allowing them to freely express their thoughts. Interviews were structured around the following discussion themes:

1. User's opinion regarding the interaction and the augmented educational experience
2. Learning environments and ages for which each system is more suitable
3. Most liked feature(s) of user's interaction with the system
4. Most disliked feature(s) of user's interaction with the system

The interview discussion followed the laddering technique [63] according to participants' answers in each one of the discussion themes.

6.3 Results

Observation during the experiment revealed certain usability problems in the interface or with the interaction technique, which were further analyzed during the interview sessions and are reflected in participants' ratings in the evaluation questionnaires. However, in certain cases, the participants—being impressed by the educational potential of the system and the novelty of the interaction techniques—tended to disregard the usability difficulties they faced and rated the system more positively than expected, based on the hypothesis that they evaluated a system prototype and that the problems will be eliminated in the final system.

In more details, SESIL was the system that was found more difficult to use, mainly due to the interaction techniques employed. Handwriting turned out to be difficult for users, it was not always accurate, and it required writing large letters. Furthermore, users faced difficulties with the pen, mainly since they found that they did not know which exact gestures were supported by the system and they were unsure of what gesture would be appropriate for each task. On the other hand, the AR study desk was considered more straightforward and easier to predict. Comments regarding the AR study desk mostly referred to enhancing its functionality with additional embedded applications (such as games). Finally, the Book of Ellie was also considered straightforward, enjoyable, and easy to predict. As shown in the diagrams below, both the AR study desk and the Book of Ellie received similar ratings from the evaluation participants.

Figure 15 displays the average and standard deviation rating for each question of the evaluation questionnaire for

all three systems. The results referring to the interaction techniques are presented in Fig. 16, where ratings regarding touch in the AR study desk and touch in the Book of Ellie have been aggregated into one result referring to touch in general. Questionnaire results related to users' characterizations of the three systems are presented in Fig. 17.

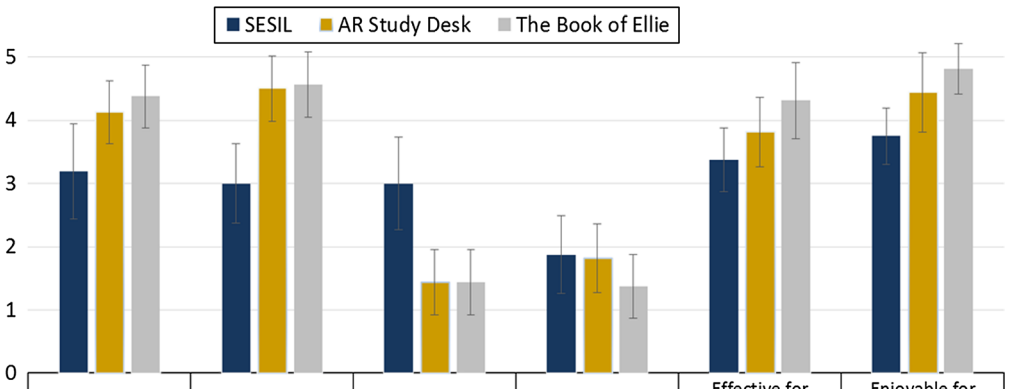
In summary, the two systems employing touch as the interaction technique were found easier to use and learn, less awkward, and obtrusive. Interestingly, however, all three systems were considered as having the potential to make studying more effective and enjoyable. The Book of Ellie was considered as the system that would make studying more effective and enjoyable, mainly since it was more targeted to specific audiences (young children), and thus, it was easier for users to appraise its learning potential.

Furthermore, results regarding the interaction techniques indicated touch- and card-based interaction as easier to use, while several problems and difficulties were encountered with handwriting- and pen-based interaction. Note that since each system employed specific interaction techniques, as illustrated in Figs. 14 and 16 includes several 0-value bars.

In addition, users considered the Book of Ellie as the most pleasant and predictable, while very close in ratings was the AR study table. On the other hand, although SESIL was considered less pleasant, straightforward and predictable, it is noteworthy that it was rated as the more interesting system.

Users' comments provided during the interview sessions were studied using affinity diagrams [64]. In summary, the following conclusions relevant to the research questions arose:

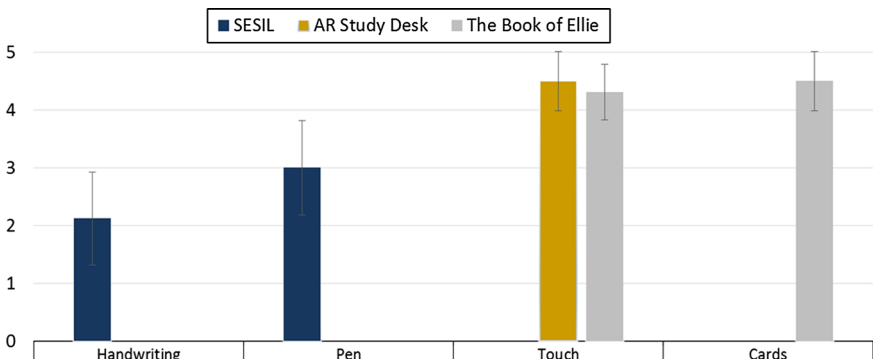
- Users' opinion regarding the unobtrusiveness of each system was greatly impacted by the ease of use. As one of the users put it regarding handwriting and SESIL: "struggling to make handwriting work disorients me seriously from my initial goal—solving a simple exercise."



	Ease of use	Learnability	Awkwardness	Obtrusiveness	Effective for studying	Enjoyable for studying
SESIL	3.1875	3	3	1.875	3.375	3.75
AR Study Desk	4.125	4.5	1.4375	1.8125	3.8125	4.4375
The Book of Ellie	4.375	4.5625	1.4375	1.375	4.3125	4.8125

Fig. 15 Analysis of the evaluation questionnaire results

Fig. 16 Analysis of the evaluation questionnaire results regarding interaction techniques



	Handwriting	Pen	Touch	Cards
SESIL	2.125	3	0	0
AR Study Desk	0	0	4.5	0
The Book of Ellie	0	0	4.3125	4.5

- The majority of users indicated that the Book of Ellie would be more appropriate for young children, while SESIL would be better to address high school students who are more familiar with handwriting. The AR study desk was considered to be a system appropriate for all ages and educational contexts, having of course its content and perhaps additional applications updated accordingly.
- Being familiar with touch-based interaction due to the popularity of smart phones and touch-enabled devices,

Fig. 17 Analysis of the evaluation questionnaire results regarding the systems' characteristics

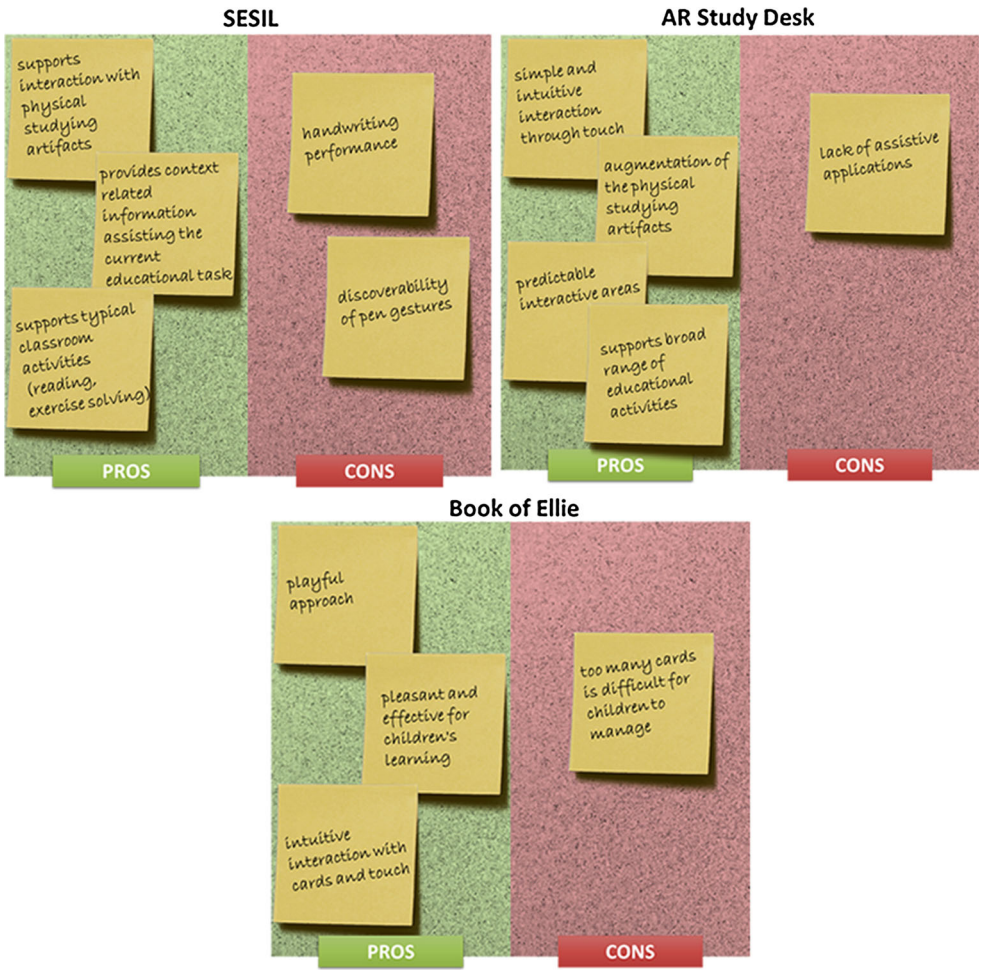


Fig. 18 Affinity diagram for users' likes and dislikes for the three systems

users found this interaction technique as the less obtrusive one. According to an evaluator's words "I don't have to think about it: I just point with my finger. It's easy and simple."

Finally, a summary of users likes and dislikes about their interaction with each one of the systems is presented in the affinity diagram of Fig. 18. Affinity diagrams were also employed in order to find out any important differences in users' comments according to their profile and mainly according to their experience with other educational software, nevertheless important differences and trends were not identified.

7 Conclusions

This paper has presented a framework aiming to enhance the educational process by augmenting physical educational assets (i.e., printed matter) with technological features in an unobtrusive and user-friendly manner. To achieve this, the framework supports natural interaction techniques and employs in the interaction typical educational material, such as books and pens. In turn, these techniques are based on visual sensing of the environment and the computational analysis of the pertinent images to extract information from it passively.

The proposed framework integrates three applications, each one structured around different educational activities and featuring different content and interaction techniques. More specifically: (1) SESIL addresses typical classroom activities such as reading and exercise-solving, (2) the AR study desk targets exploratory educational activities where the learner aims to receive information about a specific topic, either alone or in collaboration with other learners, while (3) the Book of Ellie mainly addresses the needs of younger children and edutainment activities, where learning is achieved through playing.

To evaluate and compare the various interaction techniques employed in the three systems and to assess the unobtrusiveness of the proposed applications and framework, an evaluation study was carried out. The results indicated that touch-based interaction was considered intuitive and easy to use, card-based interaction was characterized as appropriate in the context in which it was proposed, while pen-based interaction was more cumbersome for the users, due to technical difficulties with the handwriting process. Furthermore, an important parameter affecting the unobtrusiveness of a given system in the educational process was the usability of the system and the straightforward nature of the interaction technique. Another evaluation conclusion was that most of the users were positive toward using technology in the context of educational activities and that the usability problems that they encountered with a system did not affect their view regarding the adoption of the system in the educational process.

Future work will mainly address the improvement of interaction as well as the broader evaluation of the framework. To improve interaction, further unobtrusiveness and ease of use are sought through interaction with additional modalities, such as speech recognition. Besides fostering more usability needs that are important in the educational process, the system could also better adapt to the activity and context that the user is focusing on at each moment in time. Broader evaluation addresses experiments in real classroom settings, aiming at wider-range collection of evaluation data in order to more precisely characterize the educational benefits of the proposed framework and compare it to more conventional uses of technology in education, such as e-learning.

Acknowledgments This work was supported by the Foundation for Research and Technology—Hellas, Institute of Computer Science (FORTH—ICS) internal RTD Programme "Ambient Intelligence and Smart Environments." The authors would also like to thank Antonios Ntelidakis for his contribution in the artwork of Fig. 6.

References

1. Weiser, M.: The computer for the 21st century. Sci. Am. **265**(3), 94–104 (1991)
2. Abowd, G.D.: What next, ubicomp? Celebrating an intellectual disappearing act. In: Proceedings of the 2012 ACM Conference on Ubiquitous Computing, pp. 31–40. ACM, New York (2012)
3. O'Hara, K., Abigail, S.: A comparison of reading paper and online documents. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, New York (1997)
4. Spencer, C.: Research on learners' preferences for reading from a printed text or from a computer screen. J Distance Educ/Revue de l'Éducation à Distance **21.1**, 33–50 (2006)
5. de Jong, M.T., Bus, A.G.: How well suited are electronic books to supporting literacy? J. Early Child. Lit. **3**(2), 147–164 (2003)
6. Satyanarayanan, M.: Pervasive computing: vision and challenges. IEEE Pers. Commun. **8**(4), 10–17 (2001)
7. Antona, M., Margetis, G., Ntoa, S., Leonidis, A., Korozi, M., Paparoulis, G., Stephanidis, C.: Ambient Intelligence in the classroom: an augmented school desk. In: Proceedings of the 2010 AHFE International Conference (3rd International Conference on Applied Human Factors and Ergonomics), pp. 17–20. Miami, FL, USA (2010)
8. Cooperstock, J.R.: The classroom of the future: enhancing education through augmented reality. In: Proceedings of HCI International 2001 Conference on Human-Computer Interaction, pp. 688–692 (2001)
9. Margetis, G., Leonidis, A., Antona, M., Stephanidis, C.: Towards ambient intelligence in the classroom. In: Universal Access in Human-Computer Interaction. Applications and Services. Springer, Berlin, pp. 577–586 (2011)
10. Schmalstieg, D., Fuhrmann, V., Hesina, G., Szalav, Z., Encarna, L.M., Gervautz, M., Purgathofer, W.: The studierstube

augmented reality project. Presence: Teleoper. Virtual Environ **11**, 33–54 (2002)
11. Kaufmann, H.: Collaborative Augmented Reality in Education. Institute of Software Technology and Interactive Systems, Vienna University of Technology, Vienna (2003)
12. Shelton, B.E., Hedley, N.R.: Using augmented reality for teaching earth-sun relationships to undergraduate geography students. In: Augmented Reality Toolkit, The First IEEE International Workshop. IEEE (2002)
13. Juan, M.C., Carrizo, M., Giménez, M., Abad, F.: Using an augmented reality game to find matching pairs, In: Proceedings of the 19th International Conference on Computer Graphics, Visualization and Computer Vision, pp. 59–66 (2011)
14. Liu, T., Tan, T., Chu, Y.: 2D barcode and augmented reality supported English learning system. In: Proceedings of the Computer and Information Science (ICIS'07), pp. 5–10 (2007)
15. Karpischek, S., Marforio, C., Godenzi, M., Heuel, S., Michahelles, F.: Mobile augmented reality to identify mountains, In: Adjunct Proceedings of the 3rd European Conference on Ambient Intelligence (AmI-09) (2009)
16. Wellner, P.: Interacting with paper on the DigitalDesk. Commun. ACM **36**(7), 87–96 (1993)
17. Kobayashi, M., Koike, H.: EnhancedDesk: Integrating paper documents and digital documents. Asia Pacific Computer Human Interaction (APCHI'98). IEEE CS, pp. 57–62 (1998)
18. Billinghurst, M., Kato, H., Poupyrev, I.: The MagicBook: moving seamlessly between reality and virtuality. IEEE Computer Graphics, pp. 6–8 (2001)
19. Anoto.: Development Guide for Service Enabled by Anoto Functionality (2002). http://www.anoto.com
20. Liao, C., Guimbreti, F., Hinckley, K., Hollan, J.: Papiercraft: a gesture-based command system for interactive paper. ACM Trans. Comput. Hum. Interact. **14**(4), 27 (2008)
21. Luff, P., Heath, C., Norrie, M., Signer, B., Herdman, P.: Only touching the surface: creating affinities between digital content and paper. In: Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work (CSCW '04), pp. 523–532. ACM, New York (2004)
22. Forsberg, A.S., LaViola, Jr. J.J., Zeleznik, R.C.: ErgoDesk: A Framework for Two- and Three-Dimensional Interaction at the ActiveDesk. Second International Immersive Projection Technology Workshop, pp. 11–12 (1998)
23. Jeong, H.T., Lee, D.W., Heo, G.S., Lee, C.H.: Live book: A mixed reality book using a projection system. In: 2012 IEEE International Conference on Consumer Electronics (ICCE), pp. 680–681 (2012)
24. Grasset, R., Billinghurst, M., Dünser, A., Seichter, H.: The Mixed Reality Book: A New Multimedia Reading Experience. San Jose, CA, USA (2007)
25. Dymetman, M., Copperman, M.: Intelligent Paper. International Conference on Electronic Publishing, Document Manipulation, and Typography, pp 392–406 (1998)
26. Smith, D.R., Munro, E.: Educational card games. Phys. Educ. **44**(5), 479 (2009)
27. Chen, P., Kuo, R., Chang, M., Heh, J.-S.: Designing a trading card game as educational reward system to improve students' learning motivations. Transactions on Edutainment III, pp. 116–128 (2009)
28. Korozi, M., Leonidis, A., Margetis, G., Koutlemanis, P., Zabulis, X., Antona, M., Stephanidis, C.: Ambient educational minigames. In: Proceedings of the International Working Conference on Advanced Visual Interfaces, pp. 802–803. ACM, New York (2012)
29. Antona, M., Leonidis, A., Margetis, G., Korozi, M., Ntoa, S., Stephanidis, C.: A student-centric intelligent classroom. In: Keyson, D.V., Maher, M.L., Streitz, N., Cheok, A. Augusto, J.C. (eds.) Proceedings of the Second international conference on Ambient Intelligence (AmI'11), pp. 248–252. Springer, Berlin (2011)
30. Leonidis, A., Korozi, M., Margetis, G., Ntoa, S., Papagiannakis, H., Antona, M., Stephanidis, C.: A glimpse into the ambient classroom. Bull. IEEE Tech. Comm. Learn. Technol. **14**(4), 3–6 (2012)
31. Rekimoto, J.: Smartskin an infrastructure for freehand manipulation on interactive surfaces. In: CHI, pp. 113–120 (2002)
32. Streitz, N., Tandler, P., Muller-Tomfelde, C., Konomi, S.: Roomware towards the next generation of human-computer interaction based on an integrated design of real and virtual worlds (2001)
33. Wilson, A.: Playanywhere: a compact interactive tabletop projection-vision system. In: UIST, pp. 83–92 (2005)
34. Han, J.: Low-cost multi-touch sensing through frustrated total internal reflection. In: UIST, pp. 115–118 (2005)
35. Gross, T., Fetter, M., Liebsch, S.: The cuetable: cooperative and competitive multi-touch interaction on a tabletop. In: CHI, pp. 3465–3470 (2008)
36. Gaver, W., Bowers, J., Boucher, A., Gellerson, H., Pennington, S., Schmidt, A., Steed, A., Villars, N., Walker, B.: The drift table: designing for ludic engagement. In: CHI, pp. 885–900 (2004)
37. Microsoft (Microsoft surface). http://www.surface.com
38. Dietz, P., Leigh, D.: Diamondtouch: a multi-user touch technology. In: UIST, pp. 219–226 (2001)
39. SMART: Smart table (2008). http://www.smarttech.com/
40. Pinhanez, C.: Using a steerable projector and a camera to transform surfaces into interactive displays. In: CHI, pp. 369–370 (2001)
41. Kjeldsen, R., Pinhanez, C., Pingali, G., Hartman, J., Levas, T., Podlaseck, M.: Interacting with steerable projected displays. In: FG (2002)
42. Jones, B., Sodhi, R., Campbell, R., Garnett, G., Bailey, B.: Build your world and play in it: Interacting with surface particles on complex objects. In: ISMAR, pp. 165–174 (2010)
43. Goldberg, D., Richardson, C.: Touch-typing with a stylus. In: Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems, pp. 80–87. ACM, New York (1993)
44. Harrison, C., Benko, H., Wilson, A.: Omnitouch wearable multitouch interaction everywhere. In: UIST, pp. 441–450 (2011)
45. Song, P., Winkler, S., Tedjokusumo, J.: A tangible game interface using projector-camera systems. In: HCI, pp. 956–965 (2007)
46. Grammenos, D., Michel, D., Zabulis, X., Argyros, A.: Paperview augmenting physical surfaces with location-aware digital information. In: TEI, pp. 57–60 (2011)
47. Reitmayr, G., Eade, E., Drummond, T.: Localisation and interaction for augmented maps. In: ISMAR, pp. 120–129 (2005)
48. Zabulis, X., Koutlemanis, P., Grammenos, D.: Augmented multitouch interaction upon a 2-DOF rotating disk. International Symposium on Visual Computing, Rethymno (2012)
49. Cook, D.J., Das, S.K.: How smart are our environments? An updated look at the state of the art. J. Pervasive Mob. Comput. **3**(2), 53–73 (2007)
50. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
51. Margetis, G., Zabulis, X., Koutlemanis, P., Antona, M., Stephanidis, C.: Augmented interaction with physical books in an Ambient Intelligence learning environment. Int. J. Multimed. Tools Appl. **67**(2), 473–495 (2013)
52. Bookstein, F.L.: Morphometric Tools for Landmark Data. Cambridge University Press, Cambridge (1991)
53. Gelb, A.: Applied Optimal Estimation. MIT Press, New York (1974)

54. Smith R., Chang S.: VisualSEEk: A Fully Automated Content-Based Image Query System. ADM Multimedia, pp. 87–89 (1996)
55. Duda, R., Hart, P.: Use of the Hough transformation to detect lines and curves in pictures. Commun. ACM **15**, 11–15 (1972)
56. Zabulis, X., Koutlemanis, P., Baltzakis, H., Grammenos, D.: Multiview 3D Pose Estimation of a Wand for Human-Computer Interaction. International Symposium on Visual Computing, Las Vegas, Nevada, USA, pp. 104–115 (2011)
57. Wilson, A.: Using a depth camera as a touch sensor. In: ACM International Conference on Interactive Tabletops and Surfaces, pp. 69–72 (2010)
58. Margetis, G., Ntelidakis, A., Zabulis, X., Ntoa, S., Koutlemanis, P., & Stephanidis, C.: Augmenting physical books towards education enhancement. In: Workshop on User-Centred Computer Vision, in Conjunction with Workshop on the Applications of Computer Vision (2013)
59. Papadaki, E., Zabulis, X., Ntoa, S., Margetis, G., Koutlemanis, P., Karamaounas, P., Stephanidis, C.: The Book of Ellie: An Interactive Book for Teaching the Alphabet to Children. IEEE International Conference on Multimedia and Expo Workshops, San Jose, California, USA (2013)
60. Nielsen, J.: Usability Engineering, pp. 165–206. Morgan Kaufmann, San Francisco (1994)
61. Nielsen, J.: Usability Engineering, pp. 207–226. Morgan Kaufmann, San Francisco (1994)
62. Lewis, C.: Using the "thinking-aloud" method in cognitive interface design. IBM TJ Watson Research Center (1982)
63. Reynolds, T.J., Gutman, J.: Laddering theory, method, analysis, and interpretation. J. Advert. Res. **28**(1), 11–31 (1988)
64. Brush, B.: Ubiquitous computing field studies. In: Krumm, J. (ed.) Ubiquitous computing fundamentals, pp. 161–202. CRC Press, New York (2009)