

A SMART ENVIRONMENT FOR AUGMENTED LEARNING THROUGH PHYSICAL BOOKS

George Margetis¹, Panagiotis Koutlemanis¹, Xenophon Zabulis¹, Margherita Antona¹ and Constantine Stephanidis^{1,2}

¹ Foundation for Research and Technology – Hellas (FORTH), Institute of Computer Science, Heraklion, Crete, GR-70013, Greece

² University of Crete, Department of Computer Science, Greece
Email: {gmarget, koutle, zabulis, antona, cs}@ics.forth.gr

ABSTRACT

Today, augmented reality is evolving towards sophisticated approaches exploiting the opportunities offered by immersive environments and high quality 3D graphics. Such systems can prove to be very useful in the context of education, especially when learning involves reading and writing activities. However, the majority of existing systems relies on touch-based interaction, or on interaction with proprietary technological artifacts. This paper presents an approach to enhancing reading and writing on physical books through smart augmentation. It is based on the intuitive and unobtrusive monitoring of students gestures during reading and writing activities through cameras, facilitating context – aware content sensitive assistance without requiring any special interaction device.

Index Terms— augmented reality, augmented books, page recognition, gesture recognition, handwriting, ambient intelligence

1. INTRODUCTION

In recent years, the emergence of Augmented Reality (AR) and Ambient Intelligence (AmI) is driving a transition from traditional human computer interaction to natural and intuitive interaction with everyday things [8]. AmI is often claimed to bring a significant potential in the domain of education [19], while Information and Communication technologies already permeate educational environments in many ways. They can play an important role in education by increasing students' access to information, enriching the learning environment, allowing students' active learning and collaboration and enhancing their motivation to learn [14].

Recognizing the importance of reading and working on physical books, research work has been conducted towards enhancing the book reading process through AR. For example, in [22] the authors developed a system which enhances a physical book, thus providing to young readers the opportunity to interact live with its content, by placing fiducial markers on it and seeing augmented 3D graphics, through a handheld display. Reading and interacting with

this augmented book provides users with a multi-sensory experience, which is difficult to achieve through other technologies.

However, work conducted so far in applications of AR techniques to physical books have focused on touch interaction, and have not taken into account hand writing. On the other hand, as it is stated in [2] “during the act of writing, there is a strong relation between the cognitive processing and the sensorimotor interaction” (p. 398), enhancing thus the students' understanding and learning ability.

This paper presents a learning support system based on augmenting the physical book and integrating handwriting. As a result, students can be further enhanced with related material when reading, and can be assisted to accomplish their writing tasks whenever appropriate. The underlying approach is based on the recognition of book pages, as well as of specific elements of interest within a page through the intuitive and unobtrusive monitoring system of students' gestures and writing, thus facilitating context – aware content sensitive assistance.

2. RELATED WORK

Since the early years of 90's the idea of digitally augmenting physical paper was intriguing enough to trigger a few first efforts in this direction. For example, DigitalDesk [21] and its successor EnhancedDesk [15] performed physical paper augmentation with technology, offering interaction with it via touching.

Since then, more sophisticated AR solutions have been proposed by exploiting the means offered by the immersive environments and high quality 3D graphics. For example, MagicBook provides the augmentation of physical books with 3D graphics and moving avatars through VR glasses, giving to the reader the sense of living pages.

The basic interaction technique in such environments is touch. Pointing and writing in augmented reality environments has also been studied, but the majority of research work is based on proprietary technological artifacts like light pens, pen with pads, haptic devices, etc. ([23], [7]).

The work presented here introduces an augmented reality environment for students' improved learning, which is based on unobtrusive monitoring of the natural reading and writing process. This environment, named SESIL, does not require any special writing device in order to monitor the student's gestures and handwriting, as it is able to perceive interaction with actual books and pens / pencils.

Furthermore, SESIL provides context – aware help to the students, launching educational applications on a near-to-the-student display. SESIL can thus enhance the learning process by unobtrusively and naturally providing additional information related to the current student's activity.

3. SMART STUDYING ENVIRONMENT

The presented system interoperates with the ClassMate smart classroom framework [1].

A functional prototype of SESIL has been deployed on the smart school desk presented in [4], which augments a standard school desk with an additional piece of furniture including a 27 inches diagonal wide touch screen and supporting vision techniques on the front part of the desk. Three front cameras were used for book and stylus posture recognition, as illustrated in Fig. 1.



Fig. 1. SESIL deployed on the augmented desk

In the following sections a detailed description of SESIL's architecture and modules is provided.

3.1. Architecture

Fig. 2 illustrates the overall architecture of SESIL. In more detail, the output of the cameras that monitor the user's book and stylus is firstly processed by the book recognition and stylus posture modules, providing thus the page identification and the exact stylus' position on the page. Also, these modules are responsible to recognize whether the stylus is writing or not. The handwriting and gestures recognition module takes as input the output of the aforementioned modules and it provides writing text and gestures recognition. It also propagates the recognized page

to the next module. Finally, the content sensitive assistance module elaborates the output of the previous modules, in order to decide the appropriate complementary information needed by the user while reading and writing on the book. This information is displayed in a near screen. For testing purposes, SESIL has been integrated in the augmented school desk described in [4], which has been provided with the required front cameras.

The next sections provide a detailed description of the SESIL's components.

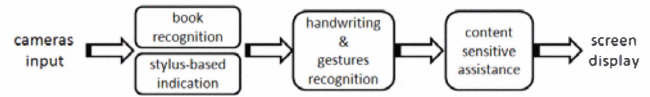


Fig. 2. SESIL's architecture

3.2. Book and page recognition

A vision process detects the presence of a known book on the desk, localizes it and recognizes the particular page that it is open at, providing a robust context cue to the application. An electronic representation of the appearance of each page is assumed. The outcome of this process is demonstrated in Fig. 3.

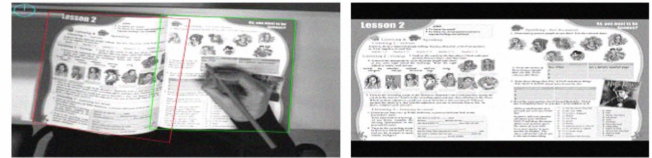


Fig. 3. Recognition of book pages. On the left, a grayscale version of the image acquired from the central camera is shown. On the right, the matched page in the electronic representation. The left image is annotated with spatial extent of each page (green & red quadrilaterals) and book orientation (cyan circle), as estimated based on the correspondences between the image and the electronic representation.

In the electronic representation, each page has been previously imaged in a planar arrangement (either scanned or synthetically generated as a printer file). An unordered representation of each page is a priori obtained by detecting SIFT features [5] in the image of each page and associating these features with it. At run time, a page of the printed copy is imaged and recognized using this representation.

When a frame is acquired, it is initially warped to the plane of the desk, providing image W, within which SIFT features are detected. These features are compared with the ones already in memory, on a per page basis. The page that provides the greatest hit ratio is considered as the matching page. To optimize the search, consecutive and preceding

pages to the current, are considered first and if they provide a very high ratio are immediately considered as matching.

The locations of the point correspondences between W and the scanned pages provide a basis for the localization of the book on the desk. For each frame, the homography that warps each imaged page (out of 2) to the scanned one is estimated through a RANSAC procedure, along with a subset of confidently correct correspondences (inliers). Using the 3×3 homography matrix, the corners of each scanned page are projected in the acquired image. These 4 projections mark the spatial extent of the book in W . Procrustes analysis is then employed [10], upon the inlying correspondences to estimate it's the 2D orientation of the book within W , or equivalently, the desk.

The process is parallelized on the GPU of the computer. At each frame that a book is found, it outputs an event with the ids of the two (or one, i.e. if the cover is matched) imaged pages along with the coordinates of each page in the desk's coordinate system.

3.3. Stylus posture

A vision process estimates the position and orientation of a colored stylus in the space above the desk, while it is manipulated by the user. To provide the estimation, the process utilizes 3 synchronous views that overlook the space desk, providing images I_i , $i=\{1,2,3\}$, at each "frame". The stylus is reconstructed as a 3D line segment, from which its location and orientation in 3D space are estimated, in 3 steps, which are all parallelized in the GPU of the computer. The outcome of this process is demonstrated in Fig. 4.

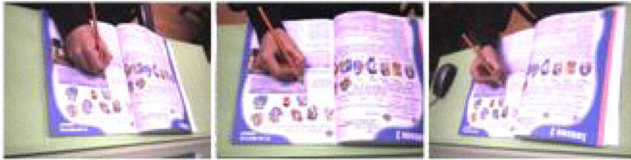


Fig. 4. Stylus pose estimation. Original images acquired from the three cameras that overlook the desk. In each image the pose estimate (location and orientation) of the stylus is superimposed as a green line segment, by projecting this estimate back to the original image.

Stylus segmentation. In each image I_i the stylus is segmented from the remainder of the image, yielding corresponding binary images M_i . Segmentation is based on a pixelwise color matching metric [24]. In M_i a pixels is 1 if matching the stylus color and 0 otherwise.

2D stylus modeling. In this step, the stylus is modeled in each M_i , as a line segment. To achieve this, a thinning process is applied to M_i , yielding image T_i . The pixels in T_i form a size-dominant line L_i , which is detected through the Hough transform [23]. Pixels along L_i that are considered to image the stylus are grouped into line segment li and its endpoints li , p_{0i} and p_{1i} , are found.

3D stylus pose estimation. Given lines L_i in each view, the stylus is reconstructed in 3D, as follows. For each view, the plane defined by L_i and K_i is considered. Thus per view-pair $j = \{1, 2, 3\}$ a 3D line C_j is obtained by intersecting the two corresponding planes. Lines C_j are combined in the output estimate of the stylus, E , as in [20]. The endpoints of the 3D segments found in images T_i , are projected and grouped onto E to find the 3D endpoints of the line segment.

At each frame upon which the stylus is found, an event containing the two endpoints p_{0i} , p_{1i} in the coordinate system of the desk is outputted.

3.4 Stylus-based indication

A third process combines book recognition and stylus pose-estimation to provide additional interaction. The 3D coordinates of the stylus are monitored and when the endpoint of the stylus is found, approximately, in contact with the book an additional event is triggered. This event contains the point Q of the book in contact with the stylus in the book's coordinate system, as well as the id of the corresponding page. In this way, the recipient of the event is able to find the indicated content in the book.

The following process is employed to estimate Q . The point where the stylus intersects this plane of the desk is found from the stylus endpoints. This point is then found if occurring within the book's, based on its estimated spatial extent on the desk (see Sec.3.3). Using the estimated homography for each page, the point of intersection is mapped to coordinates on the electronic representation of the page.

The outcome of this process is demonstrated in Fig. 5.

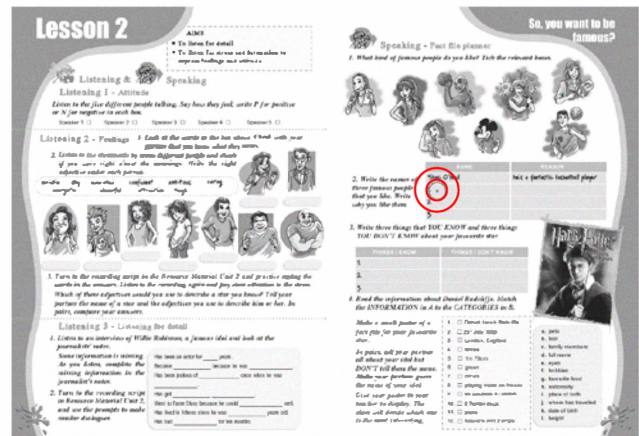


Fig. 5. Stylus-based indication of a point on the book, applied to the example of Fig. 3. The figure shows the electronic representation of the recognized page and the estimated point of stylus contact with superimposed red circles.

3.5. Handwriting and gestures recognition

This module takes as input the feedback provided by the book recognition and stylus posture and indication modules 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.

For handwriting recognition, Microsoft's Ink Analysis API [17] is used. The Microsoft Ink Analysis is an integral part of Windows 7 Tablet PC edition, and provides highly accurate handwriting and writing gestures recognition. The NHuspell spell checker [16] is also used with OpenOffice's English dictionary [18] for better recognition of the written words.

The recognized words or gestures, along with their position on the book's page and their size are fed to the next SESIL module (Content sensitive assistance). Furthermore, the recognized page is also propagated.

3.5. Book modeling and content sensitive assistance

Every book that is included in SESIL's recognition library has been stored in digital form (Portable Digital Format - PDF) and it has also been enhanced with educational meta-data. In more detail, every page of a recognizable book is classified according to the educational material 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, SESIL is able to provide in real time hints and explanations about a specific section of the exercise, which is in the user's current attention area.

```
<Page id="LE_025036_11">
  <ImageSource>file:///Path/to/pages/images/025-036_Page_11.png</ImageSource>
  <Hotspots>
    <HotSpotElement>
      <BoundingPoints>
        <Point>
          <X>0.290</X>
          <Y>0.168</Y>
        </Point>
        <Point>
          <X>0.360</X>
          <Y>0.168</Y>
        </Point>
        <Point>
          <X>0.360</X>
          <Y>0.184</Y>
        </Point>
        <Point>
          <X>0.290</X>
          <Y>0.184</Y>
        </Point>
      </BoundingPoints>
      <AssetType>FREE_TEXT</AssetType>
    </HotSpotElement>
  </Hotspots>
</Page>
```

Fig. 6. SESIL page metadata information.

The classification of books' pages is contained in a XML description file for every page, in which the interactive area

(hotspots) and their types are defined. Fig. 6 depicts the metadata file of a SESIL's page that contains a free text field of a fill-the-gap exercise. As it can be observed, every page is referenced by a unique id and it is accompanied by its digital image. This image can be displayed on any interactive screen near the physical book, enabling the user to interact with the hotspots of the digital form of the page. Furthermore, every page can contain a number of interactive hotspots denoting its interesting areas. Every hotspot 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 hotspot's bounding rectangle and the educational asset type that this hotspot denotes.

Every asset type is assigned a number of accepted stylus gestures (e.g., circle, underline, etc.) and/or writing text. In particular, the assigned writing text to a hotspot can be either a closed set of words or any word that is contained in the English vocabulary that is used, according to the asset type.

Anytime a page hotspot 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, it selects the appropriate support applications and displays them to the near screen.

The overall decision making process for choosing the suitable applications and displaying the appropriate helping content is executed by the ClassMate services with which SESIL interoperates.

ClassMate aims to provide a robust and open ubiquitous computing framework suitable for a school environment that: (i) provides a context aware classroom orchestration based on information coming from the ambient environment, (ii) addresses heterogeneous interoperability of AmI services and devices, (iii), facilitates synchronous and asynchronous communication, (iv) supports user profiling and behavioral patterns discovery and (v) encapsulates content classification and supports content discovery and filtering. Complementary to ClassMate's core an application library exists in order for other application frameworks, like SESIL, to interoperate and exploit its enhanced educative facilities, as explained in the next example.

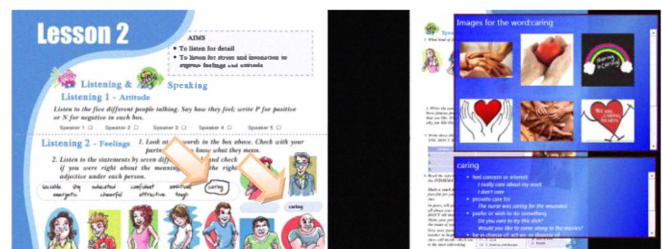


Fig. 7. SESIL screenshot example.

Fig. 7 illustrates a screenshot from a demo application which exemplifies how SESIL can assist a user trying to

accomplish a multiple-choice exercise. At the left part of the screen the image of the active page is depicted, containing the handwritten input of the user (the circle around the word “caring”). When the user circled the word “caring” in the actual book page, SESIL was activated recognizing the circle gesture around the hotspot of the page, which denotes that the word “caring” can be one of the user’s choices. SESIL also added the selected word under the corresponding avatar of the exercise and launched two informative applications: a) an images’ collection related to the word “caring” and b) definition, examples and synonyms of the word “caring”, helping the user to strengthen his confidence regarding the correctness of his choice and allowing him to retrieve / revisit related educational information. In order for SESIL to decide the corresponding result it should present to the user (e.g. in the above example, appropriate context aware help is provided when a student circles a word), it interoperates with ClassMate’s Context Manager module that directs it to consult the User Profile module and eventually retrieve the appropriate educational assets stored in the Data Space module in the form of LOM objects.

4. RESULTS

The system was tested to recognize book pages with a 100% confidence in books of approximately 200 pages, in a variety of orientation and occlusion conditions. The maximum relative orientation was 60 degrees while the worst case of occlusions covered ~50% of a page. In practice, it was observed that approximately a third of each page must be visible to be recognized. The accuracy of the stylus point of contact was always below 1cm, thereby facilitating the robust stylus-based indication of paragraphs, figures, equations, etc., within the book.

Following a user-centred design approach, and since a functional yet early prototype of SESIL was available, a formative usability evaluation was carried out by usability and interaction design experts, aiming to assess the usability and usefulness of the system. The evaluation was carried out by four experts, who were asked to carry out a heuristic evaluation [13] in order to identify potential usability problems. After interacting with the system, the evaluators were asked to fill-in a short questionnaire in order to assess conformance with the five dimensions of usability [25] and the anticipated overall user experience [12], [9].

The heuristic evaluation resulted in 7 usability improvement suggestions, 2 of which were classified as major problems, 3 as minor problems and 2 as aesthetic problems only. The evaluation suggested that the dictionary and images application could be enhanced with further functionality, allowing for example users to add a word to their personal vocabulary.

The questionnaire given to experts in order to assess the usability and overall anticipated user experience comprised of thirteen questions, providing pairs of contrasting

characterizations for the system. The answers were analyzed individually (see Fig. 8), where the rating for each attribute ranges from 1 to 7 (1 represents the most positive attitude, e.g., pleasant, and 7 the most negative attitude, e.g., unpleasant). Answers were also grouped in three categories, namely usability, learnability and user experience, with each one of the categories rated on a scale from 0 to 10. Usability received an average score of 8.21 (stdev: 0.28), learnability scored 8.09 in average (stdev: 0.3) and user experience was rated with 7,5 (stdev: 0.27). These results can only be used as qualitative measures, while more detailed quantitative results will be obtained in future evaluations with users.

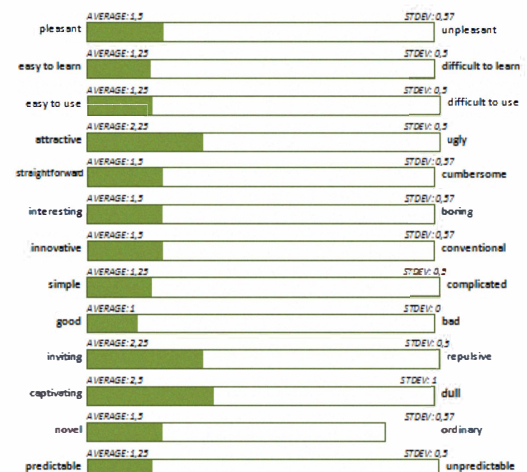


Fig. 8 Questionnaires results

5. CONCLUSION

This paper has presented SESIL, a system targeted to enhance 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. SESIL innovatively combines book and page recognition, stylus position identification, and handwriting recognition without requiring any special device apart from cameras. Additionally, SESIL exploits educational metadata on the book’s content to decide at run-time the type of additional information and support need to be provided in a context-dependent fashion.

SESIL seamlessly combines reading and handwriting on physical books with the provision of related digital content through natural interaction. The performed tests have confirmed that SESIL is robust and reliable enough for practical use, and formative usability evaluation has yielded positive results. Future work will focus on user-based evaluation in the context of a (simulated) AmI classroom environment, also assessing the educational benefits and implications of the overall approach.

6. ACKNOWLEDGMENTS

This work is supported by the FORTH-ICS internal RTD Programme 'Ambient Intelligence and Smart Environments'. The authors would like to thank Mrs. Stavroula Ntoa for her contribution to the usability evaluation of SESIL.

7. REFERENCES

- [1] A. Leonidis, G. Margetis, M. Antona, C. Stephanidis, "ClassMATE: Enabling Ambient Intelligence in the Classroom", *World Academy of Science, Engineering and Technology*, issue 66, pp. 594 – 598, retrieved 16 February 2011 from: <http://www.waset.org/journals/waset/v66/v66-96.pdf>
- [2] A. Mangel, J.-L. Velay, "Digitizing literacy: reflections on the haptics of writing", in Zadeh M.H. (Ed.), *Advances in Haptics*, pp. 385–401, 2009.
- [3] A. S. Forsberg, J. J. LaViola Jr., R. C. Zeleznik, "ErgoDesk: A Framework for Two- and Three-Dimensional Interaction at the ActiveDesk", *Second International Immersive Projection Technology Workshop*, pp. 11-12, 1998.
- [4] Antona, M., Margetis, G., Ntoa, S., Leonidis, A., Korozi, M., Paparoulis, G. & Stephanidis, C. (2010). Ambient Intelligence in the classroom: an augmented school desk. In W. Karwowski & G. Salvendy (Eds.), *Proceedings of the 2010 AHFE International Conference (3rd International Conference on Applied Human Factors and Ergonomics)*, Miami, Florida, USA, 17-20 July. [CD-ROM].
- [5] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", *International Journal of Computer Vision*, 60, 2, pp. 91-110, 2004.
- [6] D. J. Cook, J. C. Augusto, V. R. Jakkula, "Ambient intelligence: Technologies, applications, and opportunities.", *Pervasive and Mobile Computing*, Volume 5, Issue 4, Pages 277-298, August 2009.
- [7] D. Schmalstieg, A. Fuhrmann, G. Hesina, Z. Szalav, L. M. Encarna, M. Gervautz, W. Purgathofer. "The studierstube augmented reality project", *Presence: Teleoper. Virtual Environ.* 11, pp. 33-54, 2002.
- [8] E. Aarts, J.L. Encarnacao, "True Visions. The Emergence of Ambient Intelligence," Springer, 2008, ISBN 978-3-540-28972-2
- [9] E. Law, V. Roto, M. Hassenzahl, A. Vermeeren, J. Kort. "Understanding, Scoping and Defining User Experience: A Survey Approach", in *Proceedings of Human Factors in Computing Systems conference, CHI'09*, April 4-9, 2009, Boston, MA, USA (2009).
- [10] F.L. Bookstein, "Morphometric tools for landmark data", Cambridge University Press, (1991).
- [11] FORTH-ICS Ambient Intelligence Programme, retrieved on 16 February 2011 from: <http://www.ics.forth.gr/ami>
- [12] ISO FDIS 9241-210:2009. Ergonomics of human system interaction - Part 210: Human-centred design for interactive systems (formerly known as 13407). International Organization for Standardization (ISO). Switzerland
- [13] J. Nielsen and R. L. Mack. "Usability Inspection Methods", John Wiley & Sons, Inc., New York, pp. 25 – 61, 1994
- [14] M. Billinghurst, H. Kato, I. Poupyrev, "The MagicBook: Moving Seamlessly between Reality and Virtuality", *IEEE Computer Graphics*, pp. 6-8, 2001.
- [15] M. Kobayashi, H. Koike, "EnhancedDesk: Integrating paper documents and digital documents", *Asia Pacific Computer Human Interaction (APCHI'98). IEEE CS*, pp. 57–62, 1998.
- [16] Maier-hofer software, NHunspell: C#.NET free Spell Checker, Hyphenator, Thesaurus. Hunspell spelling corrector and hyphenation for C# and Visual Basic, retrieved on 16 February 2011 from: <http://nhunspell.sourceforge.net/>
- [17] Microsoft, Ink Analysis Overview (Windows), retrieved on 16 February 2001 from: [http://msdn.microsoft.com/en-us/library/ms704040\(v=VS.85\).aspx](http://msdn.microsoft.com/en-us/library/ms704040(v=VS.85).aspx)
- [18] Open Office Wiki, Dictionaries, retrieved on 16 February 2011 from: <http://wiki.services.openoffice.org/wiki/Dictionaries>
- [19] P. C. Abrami, R. M. Bernard, C. A. Wade, R. F. Schmid, E. Borokhovski, R. Tamim, M. Surkes, G. Lowerison, D. Zhang, I. Nicolaidou, S. Newman, L. Wozney, A. Peretiatkiewicz, "A Review of E-learning in Canada: A Rough Sketch of the Evidence, Gaps and Promising Directions", *Canadian Journal of Learning and Technology*, v.32 (3), Fall / automne, 2006. Retrieved 16 February 2011 from: <http://www.cjlt.ca/index.php/cjlt/article/view/27/25>
- [20] P. H. S. Torr and A. Zisserman. "Robust parameterization and computation of the trifocal tensor", *Image and Vision Computing*, vol. 15, pp. 591-605, 1997.
- [21] P. Wellner, "Interacting with paper on the DigitalDesk", *Commun. ACM* 36, 7 (July), pp. 87–96, 1993.
- [22] R. Grasset, A. Duenser, H. Seichter, M. Billinghurst, "The mixed reality book: a new multimedia reading experience.", *CHI '07 extended abstracts on Human factors in computing systems*, pp. 1953 – 1958, 2007.
- [23] R. O. Duda, and P. E. Hart, "Use of the Hough Transformation to Detect Lines and Curves in Pictures," *Comm. ACM*, Vol. 15, pp. 11–15 (January, 1972)
- [24] R. Smith and S. Chang. VisualSEEK: a fully automated content-based image query system. In *ADM Multimedia*, pages 87-89, 1996.
- [25] W. Quensbery. "The five dimensions of usability", in M. J. Albers & B. Mazur (Eds.) *Content and Complexity*, Routledge, pp. 75 – 94, 2003.