# Computer Vision Tools for eLearning

Flávio L. Coutinho, Thiago T. Santos, and Carlos H. Morimoto *

LaTIn - Laboratório de Tecnologias para Interação
Universidade de São Paulo
São Paulo, SP 05509 Brazil
{flc,thsant,hitoshi}@ime.usp.br

**Abstract.** This paper describes an academic research project in computer vision and distance learning currently under development in Brazil, and it can be roughly classified in the 1st workshop area, about computer vision research that addresses educational problems in developing regions. In particular, we describe our experience in developing two computer vision tools for distance learning (eLearning) applications. The first tool is a low cost eye-gaze tracker (EGT) integrated with an intelligent tutoring system (ITS). Due to the elevated cost of commercial EGTs, a low cost EGT might help researchers from several different areas such as psychology and human computer interaction, particularly in developing countries where resources are limited. The second tool is an extensible video browsing framework to allow exploration in rich media content. These tools are being developed to be part of the TIDIA-Ae platform. The project is funded by the *Fundação de Amparo à Pesquisa do Estado de São Paulo* (FAPESP - the state of São Paulo research funding agency), with the purpose of providing a common open source eLearning platform to all educational institutions in the state of São Paulo. By sharing our experience in this workshop we hope to contribute with the identification of common problems faced by other computer vision research labs in developing countries, discuss possible solutions, and find potential future partners.

## 1 Introduction - Education in Brazil

During the "Quality Education for All" session at the 2007 World Economic Forum on Latin America the low quality of education in Latin America was identified as one of the key factors responsible for retarding its global competitiveness. The Inter-American Dialogue "Report Card on Latin American Education" gave the region a "D", or poor grade, for its test scores, poor school conditions, national standard setting, and efforts to improve teacher quality.

In Brazil, according to its 2000 census data, about 12.9% of the population (15.5 million people) age 15 and older are not able to understand what they read. This amount is about twice the number of people with a college degree in

* We would like to thank the Fundaćão de Amparo à Pesquisa do Estado de São Paulo (FAPESP) for their financial support.

Brazil. With poor educational quality and low number of scientists and engineers the ability to innovate required to compete in a "globalized" world is crippled. According to the 2006-2007 Global Competitiveness Report from the World Economic Forum [5], Brazil has decreased its rank from 57th in 2005 to 66th place in 2007. When compared to the developing countries in the group known as BRIC (Brazil, Russia, India, and China), Brazil now falls behind in 4th place. The future is also not bright if one considers the over a decade long political-economical focus on inflation control with not enough investments in public education, and a decrease in investments in Science and Technology during the period of 2000-2005, from 1.43% of the country's GDP in 2000 to 1.36% in 2006 [4].

Although school attendance have been raising due to governmental incentives given to low income families to keep their children in school, no incentives are given to teachers to improve the educational quality. On the contrary, many teachers are forced to teach several classes (sometimes overcrowded) to compensate for low wages, quite often in different schools. Therefore they have little time for grading tests and papers, preparing their classes, updating their teaching skills, or recycling their educational material.

Technology can have a deep impact in the quality of education by helping teachers and students of all levels alike. In this paper we describe our participation in the TIDIA-Ae consortium, a collaborative effort of several educational institutions from the state of São Paulo to build a common eLearning platform. Our participation focus on the development of computer vision tools such as a low cost eye-gaze tracker integrated with an intelligent tutoring system (ITS), and a video browsing framework to facilitate data exploration in rich media databases.

The TIDIA-Ae Project (*Tecnologia da Informação para o Desenvolvimento da Internet Avançada - Aprendizagem Eletrônica* or Information Tecnology for Development of Advanced Internet - eLearning), targets both research and development in the area of Distance Learning supported by high-speed networks. The objectives include the specification, design and implementation of a comprehensive set of tools for eLearning built as low cost flexible solutions that should have profound social impact as a result of being built as Open Source Software tools that can be combined and extended as needed.

In the next section the computer vision tools under development for the TIDIA-Ae platform are described. The low cost eye-gaze tracker is introduced in Section 2.1, and in Section 2.2 we present the gaze enhanced ITS, which is being built using the GInX architecture defined in Section 2.3. Section 2.4 presents the video browsing tool that will also be integrated in the TIDIA-Ae platform. In Section 3 we describe our experience in developing computer vision research in Brazil. We start with a brief description of our sources of funding, how this particular research is being developed in collaboration with other Brazilian institutions, and then the major difficulties to conduct this project. Section 4 concludes the paper.

## 2 Computer Vision Tools for eLearning

The Laboratory of Technologies for Interaction (LaTIn) of the Computer Science Department of the Institute of Mathematics and Statistics (IME) of the University of São Paulo (USP), was created in 1999 to research and develop new technologies to enhance the user experience using computers. From 10/1999 to 03/2005, we have collaborated with the IBM Almaden Research Center, in San Jose, California, in the development of attentive interfaces. As a result from this collaboration, some efficient computer vision techniques to observe the user's actions (eye gaze tracking, face detection and tracking, gesture recognition, etc) were developed. Such techniques have been successfully applied to enhance current graphical user interfaces, and created some alternatives to the way we interact with computers today.

### 2.1 Low Cost Eye-Gaze Tracker

A direct result of our collaboration with IBM was the development of a low cost eye-gaze tracker. An eye-gaze tracker (EGT) is a device that measures the position and orientation of the eye. As described in [8], remote EGT based on computer vision techniques are more appropriate for interactive applications because they offer comfort and reasonable accuracy (about $1^o$ of visual angle) when compared to non-camera based techniques. Unfortunately, the price of commercial EGT's are prohibitive for educational applications, in particular for developing countries.

Traditional remote EGT techniques requires frequent calibration because they are sensitive to head motion. A solution to this problem was proposed by Yoo *et al.* In [14] they assume that the cornea surface is a plane and describe a multiple light source technique that uses the cross-ratio invariance property over perspective transformations to create a calibration free eye-gaze tracking technique.

Figure 1 illustrates this technique. Let $(A, B, C, D)$ be one of the four LEDs attached to the corners of the monitor screen. These points are projected to the points $(A', B', C', D')$ on the cornea surface. $P$ is the center of the pupil when looking at an arbitrary point $G$ on the monitor. Assuming that $P$ is coplanar with $(A', B', C', D')$, it is possible to compute $G$ using the cross-ratio invariant property.

Assuming a typical user position and motion in front of the computer screen, Yoo and Chung [13] have demonstrated that the image position of the corneal reflections can be approximated from the image positions of the real projected reflections using a simple linear formula with a single parameter that can be estimated by a one time calibration procedure. With this new technique, Yoo and Chung were able to improve the accuracy of the system from $2^o$ to about $1^o$ of visual angle.

In [3] we show that any calibration free eye-gaze tracking technique needs to consider the actual line of gaze instead of the optical axis as defined by the center of the cornea and the pupil center. The actual line of gaze is defined by

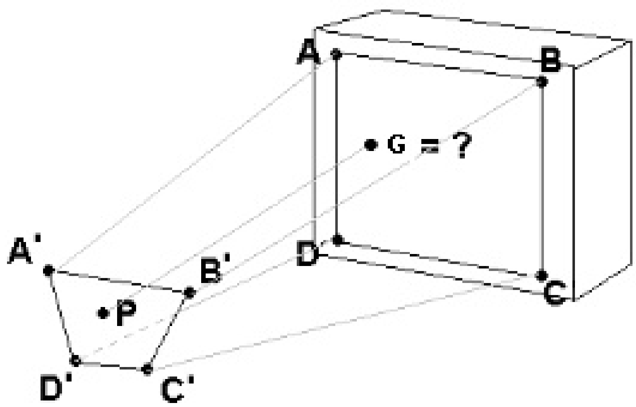


**Fig. 1.** ***Geometry of the system [3]:*** *Four LEDs are placed on the corners of the computer screen. The reflections of such LEDs on the cornea are assumed to be coplanar, and this virtual plane, as seen by the camera, is used to compute the line of gaze that connects the points $P$ and $G$.*

the center of the pupil and the center of the fovea (the region on the retina that concentrates most the color photo receptors), and in general it is about 5% off the eye optical axis, as seen in Figure 2. Because this angle is different for each eye, we have extended Yoo and Chung's method using a different calibration procedure that considers the curvature of the cornea and the true line of gaze. This calibration procedure requires the user to look at 9 points on the screen distributed along a regular rectangular grid, and only needs to be performed once per user. More details about this method are given in [3].

Figure 3 shows our current EGT prototype. The system was developed for Linux and process 30 frames per second on a 1.5GHz desktop computer. It uses a single NTSC micro camera that is synchronized with the LED's using a simple custom hardware. The synchronization is required to facilitate the segmentation of the pupil, using a differential lighting scheme as proposed in [7]. The overall extra hardware costs less than US$200,00, which is considerably cheaper than commercial EGTs. We are studying the possibility of using regular USB or firewire webcams to reduce the cost even further by eliminating the video acquisition board. Besides the cost, the tolerance to large head motion and a one time calibration per user are very important characteristics to allow users with little computer experience to access the system.

### 2.2 Gaze enhanced Intelligent Tutoring System

An Intelligent Tutoring System (ITS) can be any computer program that contains some intelligence and can be used in learning. The traditional ITS model

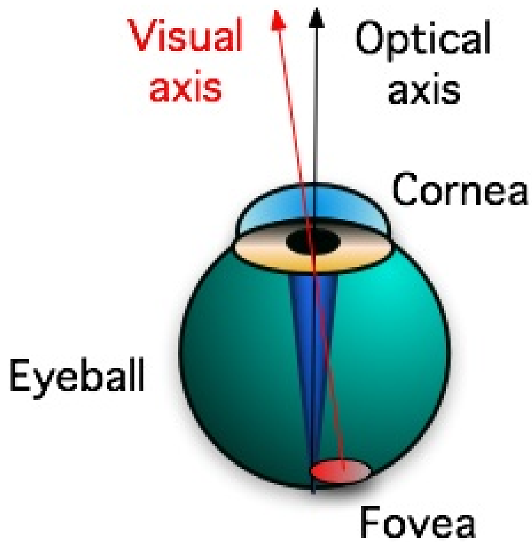


**Fig. 2.** Eye structure. Observe the difference between the optical axis, defined by the center of the pupil and the center of the eye ball, and the visual axis, defined as the center of the pupil and the center of the fovea.

contains four components: the domain model, the student model, the teaching model, and the user interface. The problems, answers, and the whole curriculum compose the domain module. The student's answer to each problem is compared to the expected answer, and the difference is used to construct a model of the student, such as his/her background and what part of the curriculum is already known or must receive more attention. The tutor module uses this information to produce the student's feedback and decide which problems have to be solved next.

The basic strategy of an ITS is to engage students in learning activities by developing their problem solving skills in a specific domain. The basic interaction paradigm is to present the student with a problem, compare its solution with the student solution, and give feedback based on the differences. This result is also used to refine the student model before repeating the process with a new problem. A new problem is selected based on what the student knows, which part of the curriculum is to be taught next, and how to present the material.

EGTs have been suggested in [6, 12] to enhance the ITS interaction paradigm. Merten and Conati [6] use gaze tracking data for on-line probabilistic user modeling. The system models user meta-cognitive behaviors, such as exploration and self-explanation, during interaction with the ITS, and uses this model to pro-

**Fig. 3.** Low cost eye-gaze tracking system.

vide adaptive support to improve these meta-cognitive behaviors and consequent student learning.

Wang *et al.* [12] describe a multiple user eLearning interface with multiple tutoring character agents that use eye-gaze information to facilitate empathy-relevant reasoning and behavior of the agents. Because gaze is highly correlated to the user's attention and interests, the agents can customize its behavior for each user, and exchange information of different students.

These two examples show that gaze data can be used to control the behavior of the tutor and the user model. Besides gaze information, affective information can also be estimated from other facial features, such as lips and eyebrows, so that the system can react not only to cognitive cues but affective cues as well [10]. In the remaining of this section we describe how our EGT is being integrated with the ITS using a learning by repetition paradigm.

### 2.3 The Gaze based Interface Extensions Architecture

The ITS is being built using the Gaze based Interface Extensions (GInX) introduced in [2]. GInX is a flexible architecture that facilitates the development of gaze enhanced attentive interfaces. An attentive interface dynamically prioritizes the information it presents to its users according to context information, such that information processing resources of both user and system are optimally distributed across a set of tasks. Vertegaal [11] suggests the classification of different types of attentive interfaces based on their ability to monitor the user's attentive state, the kind of measurements used by the sensing technology, and the way in which an attentive system may increase or decrease the load of the user, system or network.

Figure 4 shows a block diagram of the GInX architecture, which is composed of 3 modules. The domain module contains information about the task being performed by the user which is application dependent. The user module keeps information about the user state, preferences and behaviors. The attentive module receives gaze data and information from the other modules and selects an appropriate action. Observe that these modules are very similar to a traditional ITS architecture, where the attentive module substitutes the tutor. We have also deliberately not included the user interface to make the architecture application independent.

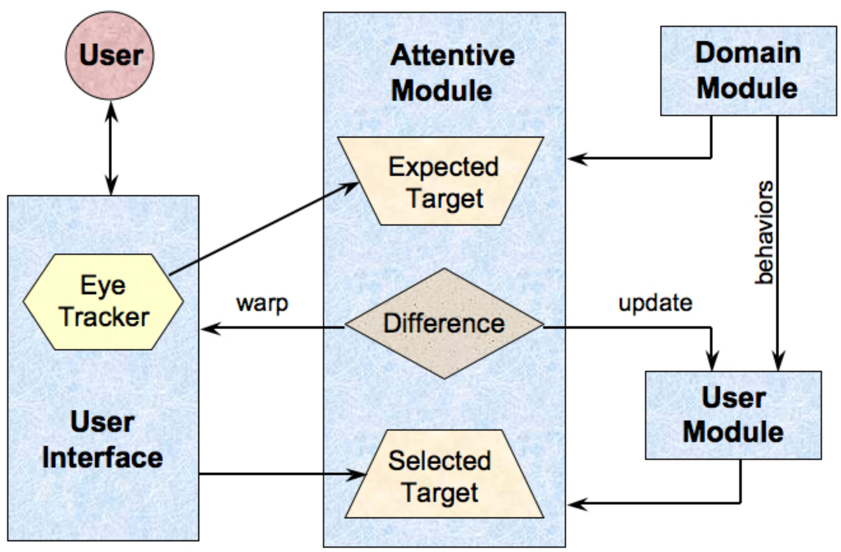


**Fig. 4.** Block diagram of the GInX architecture.

In [2] GInX was used to implement and compare the performance of 3 different pointing styles that use gaze information to directly control the cursor or to enhance the performance of regular pointing devices such as the mouse.

To extend the GInX architecture for eLearning applications the attentive module collects the history of gaze positions to compose eye behaviors. Knowledge of the widgets that compose the application interface can be used to translate gaze positions to interface targets.

The basic strategy under development is learning by repetition. A problem stored in the domain module contains information about a sequence of actions to perform a task, that can be showed to the learner during the teaching phase.

Gaze data can be used during the teaching phase to control the expected answer time of the ITS before it considers an incorrect answer and to estimate the attention levels at important targets along the presentation.

During the evaluation phase, where the student have to repeat parts of the lesson or presentation, the performance of the student can be evaluated even when his/her answer is correct using the students gaze behavior and time to complete the task. For this, the domain module must contain the expected eye behavior for each task, which may be compared to the actual student's eye behavior. When these behaviors are considerably different, this might indicate that the student had difficulty performing the lesson and the system might increase the probability of repeating it in the future. One important characteristic of gaze enhanced ITS, because the system is constantly monitoring the learner's attention, the ITS can act to redirect the learner's attention to the expected gaze behavior when the ITS identifies unexpected behavior patterns.

Because the learning activity is a repetition task, it is relatively easy to identify mistakes and to suggest ways to correct them, though system feedback must be tailored to each task. This does not imply that the learner must always mimic the exact same sequence of actions presented during the teaching phase to complete the task, since there might be more than one way to accomplish the task. A behavior may be represented, for example, as a directed cyclic graph (DAG) that represents all possible alternative patterns.

In designing computer interfaces it is common to separate tasks into 2 categories, primary and secondary tasks. Primary tasks are main tasks, such as writing a letter or report. Secondary tasks can be related to the tools required to accomplish the main task, such as using a text editor in a computer. Though quality education will be accomplished when the ITS is targeted to primary tasks, at this moment we are creating an ITS enhanced with gaze information in the form of an attentive tool that has information about the tool itself but not about the primary task. In the TIDIA-Ae project we are starting to collaborate with other groups in order to focus on learning activities, strategies and remote learning tools targeted to developing fundamental skills in sciences, in particular Calculus and Computer Science.

### 2.4 A Video Exploration Tool for eLearning

Our group is also developing a framework for an extensible video browsing tool that can be used to explore video repositories based on visual data. Its current interface is shown in Figure 5. Videos are first automatically segmented into shots using a color histogram based algorithm [9], and key frames from each shot are used to index the video. Browsing and video exploration are enhanced with filters that detect relevant objects in image frames or video shots containing that specific visual content.

The current interface allows browsing of a single video but was developed to be easily extended for small video repositories [1], for example, for course materials where pictures and videos about a particular subject can be stored without editing. The video browsing tool would allow teachers and students to easily explore the material.

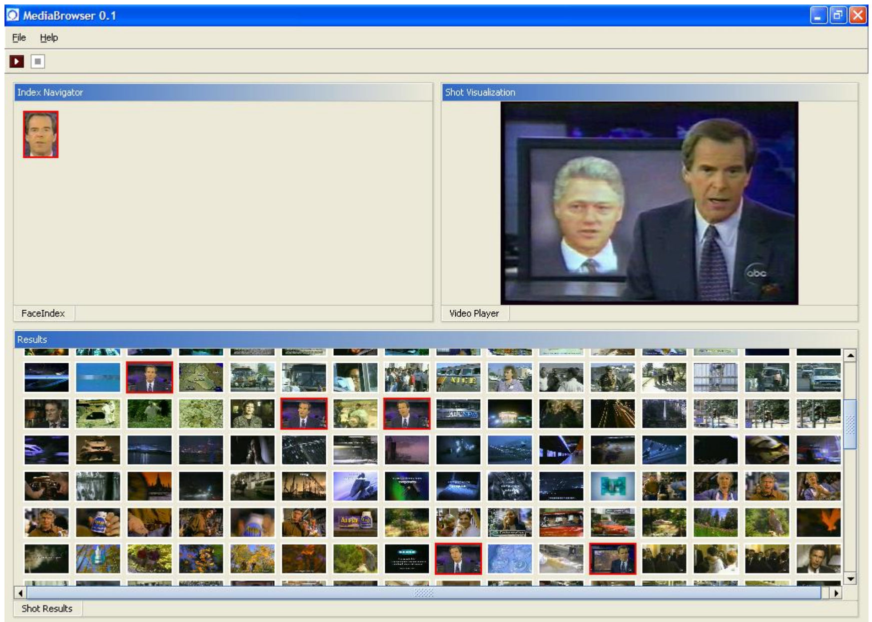


**Fig. 5.** Screen shot of the system's interface.

The concept of the tool is based on a content index navigator. The shot index is build automatically so that each shot can be accessed and viewed independently in the player subwindow. Each shot's first frame is mapped to a thumbnail that is displayed in the lower panel. The user can click the thumbnail to play the shot.

Thumbnails of relevant video objects that can be indexed automatically using filters or manually using annotation tools are displayed in the left column of the Figure 5. By clicking on one such object, the thumbnails are automatically reorganized. At the moment we have only implemented a frontal face detection filter and integrated it with PCA face recognition technique. Our experimental results based on real video data where the system is required to find the shots where a particular anchor person appears achieved a success rate of 90%.

The framework facilitates system expansion by simply adding other filters when available, such as new face models or objects. To process complex queries, the filters can be combined by means of a visual language that will allow to search for pictures/shots that contain a user selection of known objects, using basic set operations such as union, intersection, and negation.

## 3 Research Experience

To understand how research is developed in Brazil we start with a brief description of how grants are obtained. Regularly, grants are awarded to develop a specific project. The project proposals must be very detailed, and there is little flexibility to change items once they are approved. For example, a project might be related to a single graduate student, so if you have several students, you need to submit one proposal per student. This is also valid for conferences, etc.

Basically, research in Brazil is funded by 3 government agencies. Two of them are federal: CNPq (*Conselho Nacional de Desenvolvimento Científico e Tecnológico* or National Council for Scientific and Technological Development); and CAPES (*Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* or Council for Improvement of Higher Education). The third source are the FAPs (State Research Foundations) in each state (in case of São Paulo, the FAPESP). CAPES main objective is to improve higher education and it helps most graduate programs by providing them with scholarships and an annual stipend that helps to maintain each program. CNPq is more research oriented and provides funding for individual research at a federal level, while the FAPs promotes research at a state level. Therefore, there is a large overhead to write several proposals every year for equipment, new students, and attend each conference. Besides, because each proposal is evaluated by our peers (blind review), every other week we have to review a different proposal.

The University of São Paulo also contributes by complementing each project with a small amount every year, that is in general used to cover the laboratory and office supplies. Although obtaining funding for equipment is relatively simple, physical space in our computer science department is so limited that most graduate students do not have office space, so that they have to work from home or stay in the library.

Another difficulty we have is in recruiting full time graduate students. Because the city of São Paulo offers a lot of job opportunities for computer engineers and the cost of living is higher than most cities in Brazil, it is hard to keep full time students with research assistanship wages, and all funding agencies require the students to be full time. With a large number of part time students, the time required to develop a thesis increases as well as the number of students that quit without a degree. In recent years we have also noticed that the number of students interested in pursuing a graduate degree has been diminishing and apparently this is a global phenomenon in computer science and engineering.

These factors help explaining why most of the students are registered in the Master's Program. Today there is approximately 150 students enrolled in the Master's Program, and only about 45 students in the Ph.D. Program of the Computer Science Department at USP. Our program is among the top 5 graduate CS programs in Brazil, with about 37 faculty members. For example, GInX was developed by a master student, who applied and obtained a FAPESP Ph.D. scholarship, but decided to accept a job offer from the private sector. Most of the ITS is being developed by master students and only part of the

EGT is being developed by a PhD student. This pattern is similar to the group of students responsible for the development of the video browsing tools.

Our group has also collaborated with international groups, such as the Center for Automation Research (CfAR) at the University of Maryland at College Park and the IBM Almaden Research Center. In particular, the EGT pupil detection and tracking mechanism was developed in collaboration with IBM, despite a few complications due to the slow analysis process by the USP legal department, that took about 10 months to approve the joint project.

We have been participating in the TIDIA-Ae project since April 2007, but our experience with eLearning platforms dates from 2000, when our group developed the Panda system, available at `http://panda.ime.usp.br`. Panda used to support several courses at USP with more than 3000 registered students, but our difficulties in supporting and continuing the system development have turned our attention to open source alternatives such as Moodle (http://www.moodle.org) and Sakai (http://sakaiproject.org). The TIDIA-Ae is formed by a very multi disciplinary group of scientists but most with an engineering/computer science background that focus on the eLearning platform requirements and development issues. Groups with a non-IT background are participating by developing distance learning strategies and activities within their own fields, that will be supported by the TIDIA-Ae platform. It is an unusual large project for brazilian standards but at the same time very exciting and a rare opportunity to collaborate with other research institutions in São Paulo.

## 4 Conclusion

In this paper we have described our experience in developing computer vision tools for the TIDIA-Ae platform, such as an extensible video browsing tool to facilitate the exploration of rich media, a low cost eye-gaze tracker, and the Gaze Interface Extensions (GInX) architecture that is being used to develop a gaze enhanced intelligent tutoring system (ITS). All the tools described in this paper for the TIDIA-Ae platform will be available soon at the TIDIA-Ae home page (*http://tidia-ae.incubadora.fapesp.br/portal*) with a GPL license.

Within our experience, we have described the main sources of funding available for Brazilian research institutions, our difficulty in recruiting full time Ph.D. students, and obtaining office and laboratory space. With so many restrictions, the key element that allow us to do research is collaboration. Few mechanisms, such as the TIDIA-Ae project, exist to create new collaborations among public (Brazilian and international) institutions but more collaborations between academic and private institutions must be stimulated in Brazil, probably with a better regulation for all participating institutions, from governmental agencies and universities to industry and other private institutions.

## References


[1] R.A. Barbosa, F. Silva, T.T. Santos, and C.H. Morimoto. An extensible automatic video browsing tool. In *Workshop Teses e Dissertaçes - Sibgrapi 05*, Natal, RN,

October 2005.

[2] T.S. Barcelos. Interfaces prestativas baseadas em visão computacional e informação de contexto. Master's thesis, Departamento de Ciência da Computação do Instituto de Matemática e Estatística, Universidade de São Paulo, Junho 2005.

[3] F. L. Coutinho and C. H. Morimoto. Free head motion eye gaze tracking using a single camera and multiple light sources. In *Sibgrapi 06: Proceedings of the 19th Brazilian Symposium on Comptuer Graphics and Image Processing*, pages 171–178, 2006.

[4] Ministério da Ciência e Tecnologia. Investimentos nacionais em ciência e tecnologia. In *http://ftp.mct.gov.br/estat/ascavpp/gti_web/investimentos.htm*, 2006.

[5] A. Lopez-Claros, M.E. Porter, X. Sala i Martin, and K. Schwab. *The Global Competitiveness Report 2006-2007, creating an improved business environment*. World Economic Forum, 2007.

[6] Christina Merten and Cristina Conati. Eye-tracking to model and adapt to user meta-cognition in intelligent learning environments. In *IUI 06: Proceedings of the 11th international conference on Intelligent user interfaces*, pages 39–46, New York, NY, USA, 2006. ACM Press.

[7] C.H. Morimoto, D. Koons, A. Amir, and M. Flickner. Pupil detection and tracking using multiple light sources. *Image and Vision Computing*, 18(4):331–336, March 2000.

[8] C.H. Morimoto and M.R.M. Mimica. Eye gaze tracking techniques for interactive applications. *Computer Vision and Image Understanding*, 98(1):4–24, Jan 2005.

[9] T.T. Santos and C.H. Morimoto. Estruturação e indexação de vídeo digital. In *Anais do Simpósio Brasileiro de Sistemas Multimídia e Web - WebMídia 2003*, Salvador, Brazil, Novembro 2003.

[10] A Sarrafzadeh, C. Fan, F. Dadgostar, S. Alexander, and C. Messom. Frown gives game away: affect sensitive systems for elementary mathematics. In *Proc. Int. Conf. on Systems, Man and Cybernetics*, volume 1, pages 13–18, 2004.

[11] R. Vertegaal. Designing attentive interfaces. In *Proc. of the Eye Tracking Research & Applications Symposium*, pages 23–30, New Orleans, LA, March 2002. ACM.

[12] Hua Wang, Jie Yang, Mark Chignell, and Mitsuru Ishizuka. Empathic multiple tutoring agents for multiple learner interface. *wi-iatw*, 0:339–342, 2006.

[13] Dong Hyun Yoo and Myung Jin Chung. A novel non-intrusive eye gaze estimation using cross-ratio under large head motion. *Comput. Vis. Image Underst.*, 98(1):25–51, 2005.

[14] Dong Hyun Yoo, Bang Rae Lee, and Myoung Jin Chung. Non-contact eye gaze tracking system by mapping of corneal reflections. In *FGR '02: Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recogniti on*, pages 101–106, Washington, DC, USA, 2002. IEEE Computer Society.