IEEE-International Conference on Signal processing, Communications and Networking
Madras Institute of Technology, Anna University Chennai India, Jan 4-6, 2008. pp1-7

Human Centered Multimedia Computing:A New Paradigm for the Design of Assistive and Rehabilitative Environments

Sethuraman Panchanathan, Sreekar Krishna, John A. Black Jr, and Vineeth Balasubramanian

Center for Cognitive Ubiquitous Computing (CUbiC)
School for Computing and Informatics,
Arizona State University
Tempe AZ 85281, USA

Abstract: **In this paper, we describe a Human Centered Multimedia Computing (HCMC) approach to the design and development of assistive and rehabilitative environments for people with disabilities. Unlike Human Computer Interaction (HCI), which was targeted at improving the usability of computing devices, HCMC addresses a higher level of functionality where machines adapt to the users resulting in elegant interactions. At the Center for Cognitive Ubiquitous Computing (CUbiC), we are engaged in a project for individuals who are blind and visually impaired, called iCARE (information technology Centric Assistive and Rehabilitative Environment). An overview of the iCARE project components is first presented followed by a detailed discussion of one example of our research effort, namely pose angle estimation of faces towards enhancing social interactions.**

Index Terms **– Human Centered Multimedia Computing, Assistive Technology for visually impaired, Pose Angle Estimation.**

I. INTRODUCTION

Human Centered Multimedia Computing (HCMC) is a paradigm that enables sophisticated interactions between humans and computing devices/environments. Unlike Human Computer Interaction (HCI), which was targeted at improving the usability of computing devices, HCMC addresses a higher level of functionality where machines adapt to the users resulting in elegant interactions. HCMC derives its understanding of human-machine interactions through studies conducted by Social and Behavioral scientists (Psychologists, Anthropologists and Sociologists), Neurologists, Neurobiologists, etc. The recent proliferation of computing devices and their ubiquitous presence in everyday life has motivated computer scientists and engineers to work in transdisciplinary teams to draw inspiration from the study of human-human interactions in order to design effective human-machine interfaces. Examples of this approach include, Usability Engineering and User Centered Design.

Social and Behavioral scientists, Neurologists and Neurobiologists study user interactions on a narrow set of tasks to understand the capability of users in delivering and perceiving information in these interactions. Such interdisciplinary studies have resulted in an accepted model of human interactions as shown in Fig. 1. At the very basic level is *sensation* - a combination of biochemical and neurological events triggered by stimulus of sensory organs. These sensory stimuli lead up to the level of *perception* where the signals are organized, analyzed and interpreted. *Cognition* is the eventual accretion of perceptual interpretations and past experiences leading up to intelligence

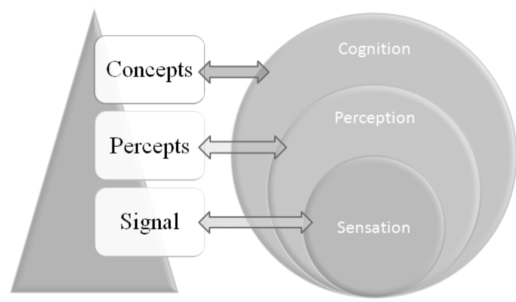

Fig. 1: Sensation-Perception-Cognition model of human interactions.

HCMC research at the Center for Cognitive Ubiquitous Computing (CUbiC) [1] is aimed at developing assistive technologies that augment the sensation, perception, and cognition of humans, regardless of the level of their natural abilities. Intensive collaborations bring together researchers from psychology, neurology, neurobiology, computer science, and electrical engineering to work on interdisciplinary HCMC research projects aimed at employing multimodal, multimedia-based technologies to augment the sensation, perception, and cognition of people ranging from highly skilled surgeons to people who are blind.

One of the major multidisciplinary research efforts in CUbiC over the last several years has been the iCARE (*information technology Centric Assistive and Rehabilitative Environments*) project. This research project is aimed at developing and applying knowledge and experience necessary toward the eventual development of assistive and rehabilitative technologies for people with a wide range of sensory, perceptive, and cognitive disabilities, including visual impairment, autism, prosopagnosia, and age-related disabilities.

978-1-4244-1923-4/08$25.00

This paper provides an overview of HCMC design approach adopted in iCARE Research activities. The paper is organized as follows. Section II presents an overarching view of four iCARE assistive technology projects, describing the motivation and purpose for developing each technology, and the role that it is intended to play in the user's day-to-day life. Section III details one of these four projects, namely iCARE-Social Interaction. Section IV concludes the paper.

II. ICARE PROJECTS FOR ASSISTING PEOPLE WHO ARE VISUALLY IMPAIRED

The iCARE assistive technology projects are aimed at helping people with visual impairments to overcome personal and social challenges that are consequences of their disability. Before embarking on the development of assistive technologies, CUbiC researchers organized focus groups of students and adults with visual impairments, their family members, their teachers, and their mobility instructors. As a result of these focus group discussions, four important challenges were identified:

1. Not being able to walk into a library, take a book off the shelf and read it.
2. Not being able to identify a person in their vicinity, until that person speaks.
3. Not being able to gain full access to vital course materials on class web sites.
4. Not being able to recognize objects at a distance, as sighted people can do.

To meet these four specific challenges, the iCARE project launched four parallel, yet interrelated, research tracks:

A. The iCARE Reader
B. The iCARE Information Assistant
C. The iCARE Haptic Interface
D. The iCARE Social Interaction Assistant

A. *The iCARE Reader*

There are millions of books in libraries around the world that will never be translated into Braille, and will never be voice recorded onto audio media. This makes them simply inaccessible to people who are blind. This constraint is felt most severely by students, and it sets limits on their educational pursuits. The iCARE Reader project is aimed at providing access to books and other educational materials to overcome this challenge. The iCARE Reader employs a camera and a portable computing device to capture images of printed material and then convert those images (in real time) into textual data, using Optical Character Recognition (OCR) technology. This textual data can then be presented interactively to the user, using Text-to-Speech (TTS) software. The functional block diagram of the Reader is shown in Fig. 2.

The iCARE Reader project is being executed in three consecutive phases: (1) a Table-top Reader, (2) a Portable Reader, and (3) a Wearable Reader. To date, the Table-top Reader and the Portable Reader have been developed and successfully deployed for evaluation by users. However, current digital camera technology is still too bulky to fit unobtrusively into a pair of eyeglasses, as planned for the wearable Reader. Fortunately, camera technology is evolving quickly, promising smaller and higher resolution cameras soon. While the desktop and portable Readers have used conventional I/O devices, such as keyboards, audio speakers and headphones, the wearable reader will demand new research efforts to develop interactive, adaptive I/O devices for closely coupling the assistive device to the human user.

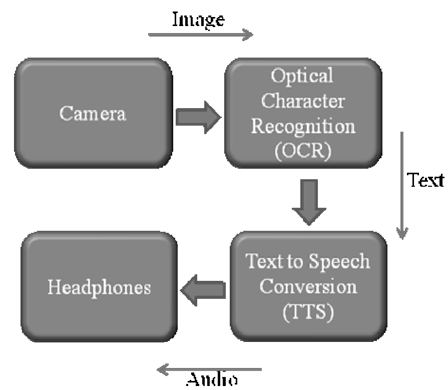

Fig. 2: Block Diagram of the iCARE Reader

B. *The iCARE Information Assistant*

Universities are relying more and more on course-specific web sites to distribute course materials, collect homework, and communicate between instructors and students. In fact, many courses are now strictly online. Most of these sites were originally designed to be accessed with a mouse. That presents little difficulty for a sighted student. However, students with visual impairments may find these sites difficult (or even impossible) to navigate.

While screen reader applications (such as Window Eyes and JAWS) provide access to textual documents through text-to-speech or dynamic Braille interfaces, they are much less effective for navigating large hypertext documents – especially those that use visual icons to represent hyperlinks. In cases where multiple links must be traversed to arrive at a desired destination, an inability to see the screen deprives the user of vital contextual information, and imposes a considerable cognitive load on the student, who might struggle to keep track of their progress during convoluted navigational processes.

The iCARE Information Assistant simplifies the navigation process by (1)crawling a class web site and parsing and categorizing various types of contextual material based on the subject being covered, and (2) reducing the cognitive load on the user by dynamically restructuring the site's content, based on the user's explicit declaration of the required task. The aim of this dynamic restructuring is to reduce the navigational distance to the content needed to complete the declared task.

The iCARE Information Assistant is built to interact with a server running the commercial online course management application called *Blackboard*, which is widely employed by universities and high schools. The research challenges faced in achieving the goal of reducing the cognitive load on users who are blind during web site navigation are as follows:

- The system needs to (1) determine what type of information the student needs, based on the user's declaration of interest, recently accessed content, current context, access history, and a student profile that contains the student's preferences and information, and (2) present the most relevant information to the student in a convenient format.
- The system needs to understand the logical relationships between the various types of content on the site, such as announcements, class schedules, homework assignments, and communications.
- The system needs to provide (1) a keystroke-based user interface that allows a student who is blind to easily navigate without a mouse, (2) guidance for learning and remembering the required keystrokes, and (3) a keyword-based search feature.

As the student uses the iCARE Information Assistant, it adapts itself to the student, while working interactively and intelligently with the Blackboard web server. This adaptive assistant allows a student who is visually impaired to work independently, saves the instructor content preparation time, and helps ensure that the student who is visually impaired has the same access to course materials as sighted students.

C. *iCARE Haptic Interface*

People who are visually impaired rely largely on their sense of hearing (audio) and touch (haptics) to perceive their environment. Hearing is the primary sensory input for environmental navigational, and touch is the primary means by which they explore and recognize objects. Recent advances in the area of haptics have revealed some surprising sensory abilities that both sighted and visually impaired people have, using their sense of touch. This research has lead to the development of novel adaptive haptic interfaces for information delivery to humans.

One of the historic difficulties in haptic perception research has been the collection of precise data on hand postures, and movements during haptic exploration. However, recent technologies have made it possible to precisely track subtle hand movements, as well as delivering vibro-tactile stimulation to the hand. Fig. 4(a) shows a haptic glove with 24 positional sensors to monitor the exact posture of hand in real time (including all joint angles) and 6 vibro-tactile stimulators that can provide real-time tactile feedback.

At CUbiC, we are exploring the use of haptic technologies, such as the glove shown in Fig. 4(a), to encode and deliver visual percepts (such as the color, texture, size, and shape of an object) to humans who are blind. As illustrated in Fig. 4(b), this would allow a user to haptically perceive the features of a distant object that is within the field of view of a wearable video camera. Haptics is a relatively unexplored research area, with many unexploited applications. A recent upsurge in interest in haptics has opened new opportunities for an expansion of CUbiC's collaborative research in the areas of the (a) the Neurology of Haptics, (b) the Psychology of Haptics, (c) Haptic Human Computer Interfaces, and (d) Multimodal Interfaces and Wearable Devices.

D. *The iCARE Social Interaction Assistant*

The ability to socially interact with other humans is a key to personal and professional success. Face-to-face interactions involve both verbal and nonverbal communications, the latter including facial expressions, eye contact, blinking, and blushing, in addition to hand and body gestures. For this reason, vision plays a crucial role in social interaction – a role that puts people with visual impairments at a disadvantage. Of particular importance to one's professional career is the need to quickly acknowledge the presence of friends and colleagues, and to initiate social interaction when encountering them unexpectedly. The iCARE Social Interaction Assistant project is aimed at the development of an assistive device that can help people who are visually impaired interact more effectively during social encounters and interactions.

Fig. 3(a) Haptic glove (b) Distal Object Perception

As shown in Fig. 3 the iCARE Social Interaction Assistant includes a tiny wearable video camera that is mounted unobtrusively on the bridge of a pair of glasses. The video stream from the camera is routed to a mobile or wearable computing device, such as a laptop or a PDA. The current implementation of the iCARE Social Interaction Assistant is able to extract faces from the video stream, and compare them to faces stored in its onboard face database which was captured with the video camera during previous conversations. When a face is recognized, the name of the recognized person is then spoken discretely to the user, along with a confidence factor. The threshold confidence factor can be adapted to the human user – some users preferring to hear a name only when the device is extremely confident, while others want to hear names even when the confidence is lower.

Some of the important research challenges being addressed in the development of the iCARE Social Interaction Assistant system include:

a) Detection and tracking of all faces within the camera's field of view, to determine the number and the locations of the people within the scene
b) Detection of significant changes within the scene, such as the entry or exit of persons
c) Recognition of the people in the scene, followed by the name of a new person entering the field of view of the camera for the first time in that context.
d) Detection of the head pose angle and the eye contact status of people within the camera's field of view
e) Recognition of significant facial expressions or facial mannerisms
f) Detection of significant hand or body gestures.

Given the flood of information that could result from all of this processing, the assistive device will need to monitor the context, as well as the state of the human user, to modulate its interaction with the user.

III. POSE ANGLE ESTIMATION

In addition to the multiple disciplines outside of engineering, the research tracks within the iCARE project span multiple research domains, including computer vision, machine learning, and pattern recognition – all of which can potentially play important roles in the development assistive technologies for people who are blind. One good example is the ability to perceive head pose angles in real time during group conversations, which presents a difficult research challenge for the iCARE Social Interaction Assistant.

Many approaches have been adopted in the past to achieve pose angle estimation. A broad subjective classification of these techniques can be found in [2]. Our approach conceptualizes a collection of images of a single face across a range of pose angles as all lying on a low-dimensional manifold within a high-dimensional feature space. To take an extreme example, a very high dimensional feature vector can be derived from each normalized grayscale face image by simply using all of the pixel values as the elements of that vector. Because there are a limited number of degrees of freedom for head movement (i.e. yaw, pitch, and roll) images of the same face at different pose angles (when projected into this very high-dimensional feature space) produce a set of vectors that define points that all lie within a low-dimensional manifold [7] [2]. For example, images of the human face with three different pose angle variations (yaw, tilt and roll) produce vectors that all terminate within a 3D manifold, which is a finite region within the higher dimensional feature space.

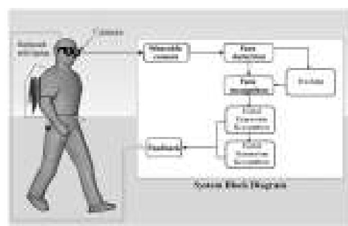

Fig. 4: Social Interaction Assistant

A. Non-linear Dimensionality Reduction using Manifold Learning

When a set of data points all lie on a manifold within the high-dimensional space, Euclidean distances across the high-dimensional space do not usefully represent the distance relationship between those points. A more useful distance measure can be derived from the pair-wise geodesic distances (which roughly define the manifold) between neighboring points. While linear techniques such as Isomap (Tenenbaum et al [9]) provide a means for estimating the distances between any pair of points across the manifold within which the points lie, Locally Linear Embedding (LLE)(Roweis and Saul [8]) and Laplacian Eigenmaps (Belkin and Niyogi [3]) adopt a local fitting approach, based on the points in the neighborhood around each point.

B. Experimental Setup

In this work, we have used the FacePix(30) database [6] [4] developed at CUbiC for our experiments and evaluation. For each of 30 people, the FacePix(30) database contains a spectrum of 181 face images that span a range from –90° to +90° of pose angle, in 1-degree increments. (A similar spectrum of 181 images is provided for a range of illumination angles.) Fig. 5 shows the apparatus that was used to capture the range of face images. A video camera and a spot light are mounted on separate annular rings, that rotate independently around a subject, who is seated in the center. Angle markings on the rings are captured simultaneously with the face images in the video sequence, which allows frames to be extracted at 1-degree increments.

The FacePix(30) database consists of two sets of face images for each person: one set with pose angle variations, and another set with illumination angle variations. Each of these image sets may be conceptualized as a 2D matrix of face images, where each matrix has 30 rows (representing the 30 different subjects), and 181 columns (representing angles from −90° to +90° at 1-degree increments). All images are 128x128 pixels, and are normalized to center the pupils vertically along a standard horizontal line position, and to scale the face size, based on the vertical distance between the eyes and the mouth. Fig. 6 shows two subsets of these face images, which includes pose angles and illumination angles ranging from −90° to +90° in steps of 10°.

Experiments were performed with a set of 24x91 = 2184 face images from the FacePix(30) database, consisting of 24 individuals with pose angles varying from -90^{o} to $+90^{o}$ in increments of 2^{o}. The images were sub-sampled to 32×32 resolution, and two different feature spaces were used to represent the content of the images. The results presented here were based on the use of a grayscale pixel intensity feature space, and a Laplacian of Gaussian (LoG) feature space.

Both of these feature extractions methods produce high-dimensional feature vectors. Various methods could be used to accomplish the nonlinear mapping from this high-dimensional feature space to the low-dimensional embedded manifold. We adopted the Generalized Regression Neural Network (GRNN) with Radial Basis Functions to learn the non-linear mapping. Once the low-dimensional embedded manifold was obtained, linear multi-variate regression was used to obtain the pose angles of each test image.

B. Comparing linear dimensionality reduction to manifold learning for pose estimation

Traditional approaches to pose estimation have used linear dimensionality techniques. However, the mapping from the original feature space to the low dimensional manifold will not, in general, be linear, so a non-linear dimensionality reduction was expected to perform better. To test this expectation, we compared the performance of the manifold learning techniques to PCA. Tables I and II show the results of our comparison. While Isomap and PCA perform similarly, both of the local approaches (i.e. Locally Linear Embedding and Laplacian Eigenmaps) consistently provided a $3\text{-}4^{o}$ better pose angle estimation than PCA.

Fig. 5: The FacePix(30) face capture platform

Fig. 6: Example image set from FacePix(30) database showing a range of pose angles and a range of illumination angles

Table 1. Results of head pose estimation using Principal Component Analysis and manifold learning techniques for dimensionality reduction, for the grayscale pixel feature space

Dimension of embedding	Error in pose estimation (low = good)			
	PCA	Isomap	LLE	Laplacian Eigenmap
10	11.37	12.61	6.60	7.72
20	9.90	11.35	6.04	6.32
40	9.39	10.98	4.91	5.08
50	8.76	10.86	4.37	4.57
75	7.83	10.67	3.86	4.17
100	7.27	10.41	3.27	3.93

To find a low-dimensional manifold that contained the face images of all 30 subjects, we proposed a supervised framework for manifold learning. The approach aligns face images on the manifold by using the known pose angles of the training images. To ensure the alignment, we proposed a Biased Manifold Embedding framework that ensures that face images whose pose angles are closer to each other are maintained nearer to each other in the low-dimensional embedding. The distances between data points in the high-dimensional feature space are biased with pose angle differences between the corresponding pose angle images of the 30 subjects. We construct a pair-wise dissimilarity/distance matrix between all 30 sets of training data points using the pose angle differences between the points. We define the modified biased distance between a pair of data points to be of the fundamental form:

Table 2. Results of head pose estimation using Principal Component Analysis and manifold learning techniques for dimensionality reduction, for the LoG feature space

Dimension of embedding	Error in pose estimation (low = good)			
	PCA	Isomap	LLE	Laplacian Eigenmap
10	9.80	9.79	7.41	7.10
20	8.86	9.21	6.71	6.94
40	8.54	8.94	5.80	5.91
50	8.03	8.76	5.23	5.23
75	7.92	8.47	4.83	4.89
100	7.78	8.23	4.31	4.52

C. Biased Manifold Embedding: A novel supervised approach to manifold learning

$D_{biased}(i, j) = \lambda_1 \times D(i, j) + \lambda_2 \times f(P(i, j)) \times g(D(i, j))$ where $D_{biased}(i, j)$ is the modified biased distance, $D(i, j)$ is the Euclidean distance between two data points x_i and x_j , $P(i, j)$ is the pose angle difference between x_i and x_j, f is any function of the pose distance, g is any function of the original distance between the data samples, and λ_1 and λ_2 are constants.

In our experiments, the biased distance between a pair of points was given by:

$$\tilde{D}(i,j) = \begin{cases} \dfrac{\propto (P(i,j))}{max_{m,n}P(m,n) - P(i,j)} D(i,j) & P(i,j) \neq 0 \\ 0 & P(i,j) = 0 \end{cases}$$

This biased distance matrix is then used for Isomap, LLE and Laplacian Eigenmaps to obtain a pose-ordered embedding.

D. Supervised manifold learning for person-independent pose estimation

Traditional manifold learning techniques produce better results for pose estimation, when compared to linear dimensionality reduction techniques. However, we hypothesize that the supervised approach to manifold learning will provide even better results for person-independent pose estimation. The proposed BME framework was applied to face images from the FacePix(30) database, and the resulting performance was compared to the performance of global (Isomap) and local (Locally Linear Embedding and Laplacian Eigenmaps) approaches to manifold learning. The results of these experiments are presented in Fig. 7 and 8.

As shown by these results, the pose estimation error significantly drops with the proposed approach. LoG is also shown to produce a better performing feature space than grayscale pixel intensities. This corroborates the intuitive assumption that the head pose is primarily represented by the large-scale features of face images, rather than the fine textural features of the images, which are represented by the pixels in the image and can be largely considered to be noise.

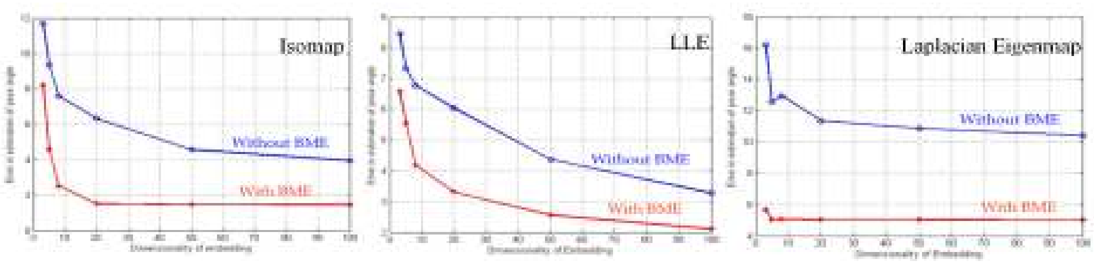

Fig. 7: Pose estimation results of the BME framework against the traditional manifold learning technique with the grayscale pixel feature space.

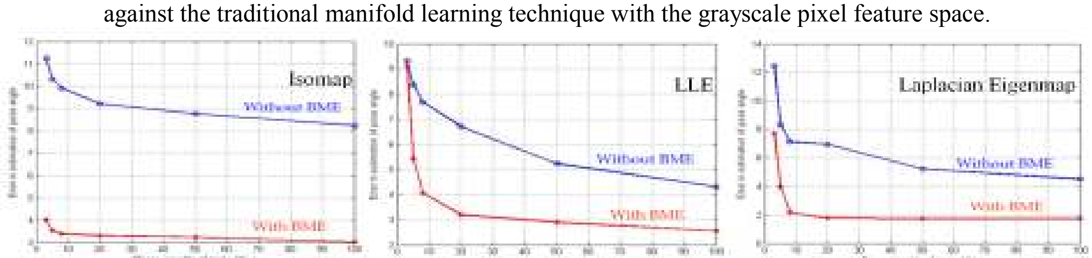

Fig. 8: Pose estimation results of the BME framework against the traditional manifold learning technique with the Laplacian of Gaussian (LoG) feature space.

These results show the merit of the proposed supervised framework for manifold learning as being more effective for head pose estimation, and we are currently incorporating this algorithm into the iCARE Social Interaction Assistant to perform real-time pose angle classification from the camera video stream. We will study potential topological instabilities that might adversely affect the performance of global manifold learning approaches such as Isomap, and will come up with solutions to circumvent such issues in pose estimation, and other face analysis applications.

IV. CONCLUSION

In this paper, we have discussed a Human Centered Multimedia Computing (HCMC) approach to the development of assistive and rehabilitative devices for people with visual impairments. The basic and applied research components of the iCARE project at CUbiC along with the details of one of the basic research issue – pose angle estimation was presented. Current and future research is focused on broadening the assistive and rehabilitative devices and environments for people with a wide range of sensory, perceptive, and cognitive disabilities, including visual impairment, autism, prosopagnosia, and age-related disabilities.

REFERENCES

[1]. http://cubic.asu.edu/

[2]. Balasubramanian V, Krishna S, Panchanathan S, "Person-Independent Head Pose Estimation using Biased Manifold Embedding", EURASIP Special Issue on Advanced Signal Processing and Pattern Recognition Methods for Biometrics, 2007.

[3]. Belkin, M. and Niyogi, P. 2003. Laplacian eigenmaps for dimensionality reduction and data representation. Neural Computation 15, 1373–1396.

[4]. Black, J., Gargesha, M., Kahol, K., Kuchi, P., and Panchanathan, S. 2002. A framework for performance evaluation of face recognition algorithms. In Proceedings of SPIE ITCom 2002 Series. Boston, Massachusetts, USA.

[5]. Herman, I., Melancon, G., Marshall, M. S. Graph visualization and navigation in information visualisation: a survey. IEEE TVCG, 6(1), 2000.

[6]. Little, G., Krishna, S., Black, J., and Panchanathan, S. 2005. A methodology for evaluating robustness of face recognition algorithms with respect to variations in pose and illumination angle. In Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing. Philadelphia, USA, 89–92.

[7]. Raytchev, B., Yoda, I., and Sakaue, K. 2004. Head pose estimation by nonlinear manifold learning. In 17th International Conference on Pattern Recognition (ICPR04). Cambridge, UK.

[8]. Roweis, S. and Saul, L. 2000. Non-linear dimensionality reduction by locally linear embedding. Science 290, 5500, 2323–2326.

[9]. Tenenbaum, J. B., Silva, V. d., and Langford, J. C. 2000. A global geometric framework for nonlinear dimensionality reduction. Science 290, 5500, 2319–2323.