Speech Based Shopping Assistance for the Blind

J. Farzana[1], Aslam Muhammad[1], A.M. Martinez-Enriquez[2], Z.S. Afraz[1], and W. Talha[1]

[1] University of Engineering and Technology, Lahore, Pakistan
farzanajbn@yahoo.com, maslam@uet.edu.pk
[2] Department of Computer Science, CINVESTAV-IPN, D.F. Mexico
ammartin@cinvestav.mx

Abstract. Vision loss is one of ultimate obstacle in the lives of blind that prevent them to perform tasks on their own and self-reliantly. The blind are trusting on others for the selection of trendy and eye-catching accessories because self –buying effort lead them in such collection that is mismatch with their personalities and society style. That is why they are bound to depend upon on their family for shopping assistance, who often may not afford quality time due to busy routine. The thought of dependency rises lack of self-confidence in blinds, absorbs their ability to negotiate, decision making power, and social activities. Via uninterrupted speech communication, our proposed talking accessories selector assistant for the blind provides quick decision support in picking the routinely wearable accessories like dress, shoes, cosmetics, according to the society drifts and events. The foremost determination of this assistance is to make the blind liberated and more assertive.

Keywords: Speech processing, image processing, knowledge based system, wearable item selection, visual impairment.

1 Introduction

According to the surveys of world Health Organization (WHO), the number of blind persons is round about 40 - 45 million. The 135 million people have low vision and approximately 314 million have some kind of visual impairment [1].The ratio of blind persons is greater in developing countries rather than industrial countries approximately 1% and 0.4% respectively [2,3]. Visual impairment can be characterized as partially or complete loss of vision, these two main categories influence the requirement of the user interfaces [4].

In general, loss of vision is a major hurdle in daily living; people have to face a series of problems in how to read, write, access information, way findings, moving liberally in fluctuating environments, selecting daily wearable items, interaction with people and surroundings. That is why everyone has his/her own set of rules and standards regarding perception. Regrettably, sometimes the way of dressing is used to signify the personality. Dressing is essential to be properly either for going to a job interview or to attend a social event, i.e., dressing provides control to represent someone as an individual in the society. Then, how can they meet such a critical criteria to represent themselves according to the standard of this society? Although the blind

J.F. Martínez-Trinidad et al. (Eds.): MCPR 2014, LNCS 8495, pp. 361–370, 2014.

persons have ability to locate day-to-day commodities by using their self-arranging methods, they cannot differentiate among products in case of colors, fashion, and new trends. That is why blind people are cautious to attend the formal functions because of the dearth of deciding on wearable items. This scarcity makes them realize their reliance on the others that results in loss of social activities and they enclosed themselves in their rooms. These emerging issues point out the need of wearable item selector assistance for the blind.

We propose a Talking Accessories Selector Assistant for the Blind (TASA-Blind) system. Our approach is based on the knowledge based systems, includes speech processing technology to provide direct communication facility and image processing heuristics to encourage the blind to express needs willingly and buy items that are well-matched to the personality at all events. Our desktop application with dual interface (speech and keyboard selection mode) is built to deploy in shopping environment where it assists customers (visually impaired and blind) as well as salesman too in understanding the needs of blinds.

Section 2 introduces the background and literature review about the speech based interface systems .Section 3 illustrates the system design and provides implementation detail of actual real system in order to validate our proposal. Finally, Section 4 summarizes our contributions and offers possible future research and expansions.

2 Related Work

Electronic Program Guide (EPG) is a multimodal media center interface, designed for sighted, visually impaired and blind persons. Thus, extended features like proper use of color contrast, zooming focus, typography, GUI coupled with speech output and haptic feedback really make the interface more useful than traditional EPG views [5]. *Framework for Blind user interface development* is useful for children in educational purposes [6].The *framework* comprises a set of user design guidelines, the programming library, and an interface development toolkit. Interface combining speech and pattern recognition enables user friendly admittance to computer [7]. After selecting one color from the color picker module, users can place it anywhere, and then shift module used to control the computer via speech commands. The *SICE framework* [8] can be used for design and development of speech based Web supporting applications. SICE provides flexibility without requiring telephony. *TrailBlazer* [9] is a CoScripter that interfaces with JAWS screen reader and facilitate users to share macro's database. The interface allows the blind to read the pseudo natural language description of each action in the macro. *Skipping and Hybrid techniques* are used in real time to process the detection of skin tone [10].Instead of testing each pixel, this technique skips predetermined number of pixels. The heuristic considers that the nearest color pixels of the skin are also skin pixels, particularly in Web mature images. *A clustering unsupervised technique* [11] is used for sorting skin color. *Step Wise Linear Discriminant Analysis* (SWLDA) technique performs forward and backward processes for image feature extraction in fast and efficient way [12]. The prominent features are selected from the feature space based on partial F-test value, and then categorize them into classes on the basis of regression values. Skin extraction using *HSV color space* extracts skin region from a given image. Firstly, the system converts image to hue, saturation, and lightness (HSV) color space and then

detects which pixels' H component lie in the range of 6 to 38[13*]. Facial and head boundary* extraction [14] is a technique of double thresholding to trace the outer (head) and inner (face) boundary in a given image having a frontal face. The input image is first smoothed using a median filter and the edges are detected using Wechsler and Kidode's method for edge detection [15].

Consequently, existing interface based applications have revolutionized the life of the blind, however selection of personal accessories remains an unresolved problem.

3 Talking Accessories Selector Assistant for the Blind: TASA-Blind

TASA-Blind is an interface based shopping system that plays the role of assistant by making assessment about wearable accessories.

a. Model Schema of TASA-Blind

The principal physical devices, responsible for user interaction with TASA-Blind are: a web cam, a microphone, a gait detector, a laser range finder device and a work station. In order to get personal information, the general architecture includes: the knowledge based system (KBS), image feature extraction process (IFE), categories description module (CDM), the speech processing module (SPM), and the graphic user interface (GUI) (see Fig. 1).

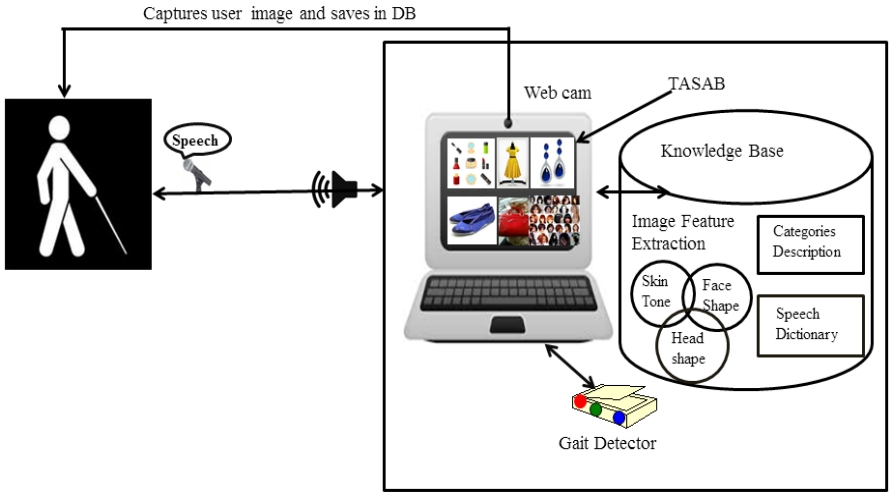

Fig. 1. Schema of the TASA-Blind system

In addition, our infrastructure includes a trial room particularly designed for the blind. All physical components of speech based assistance are equipped into the trial room, where the blind can responsively interact with the whole system in loneliness. The user's snapshot is captured, and forwarded to the IFE module (see Fig. 2). Based on voice instruction, GUI explains the way in which users can benefit of the whole infrastructure (see Fig. 2).

b. System Architecture

The information collector system extracts facial features and précises skin tone by using sampling mechanism [16]. A gait detector sensor determines the foot size and the walking style of the user [17]. This device contains pressure sensors, placed on solid surface. In order to determine user's foot size, user has to walk over the platform with bare footed. Once the user starts walking, the pressure sensors of the gait detector measure the whole user's pressure foot of different positions, the size, and flat foot, among other useful information. To measure the precise height of a person, a laser range finder is affixed in the ceiling of trail room that emits laser beam on the beneath standing person and measures height accurately [18]. Speaker and microphone are resources for the interaction between the system and the user. During all the process, users are guiding with precise verbal instructions regarding the use of the whole infrastructure. System inquires about personal information, sort of category and function, favorite item stuff, price range, etc.

1. Perception Vector (PV)

Information from sensors (image, data) and user's voice are received by PV that behaves like an interactive intermediate module with other component of TASA-Blind. PV is made up of two components: 1. Momentary storage I/O buffers for Input perception: voice, image, sensory input and output. Buffer stores input data from location sensors and from user, preserving Output data from convertor selector respectively. 2. English speech synthesizer system converts text to speech with artificial human voice [19]. Speech synthesizer has its own catalogue that is used for the storage of text description, speech dictionary. The system recognizes the user's vocabulary using the building grammar.

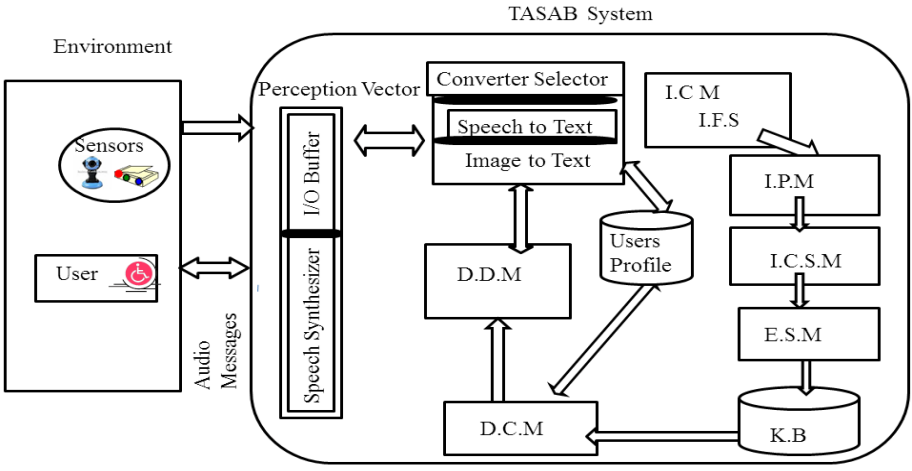

Fig. 2. Architecture of the TASA-Blind system

2. Convertor Selection Module (CSM)
Convertor selector Module (CSM) receives acuity input from PV that is being present in the I/O buffers. CSM performs a list of tasks after getting the input. The foremost task is to save a copy of user's captured image into IDB repository, then check the input format that can be either an image or speech. According to the input format, CSM directs data converting: text2speech or image2text.Existing models and algorithms are used for the conversion: Hamming Trace Transform (HTT) method for image2text, for extracting user's facial features [19]. The HTT method is blend of different concepts like Trace transform and Fourier transform, for detecting facial feature from the target image. In order to measure the skin tone, the sampling mechanism [16] is practiced, while for object detection machine learning approach is used. Machine learning methodology further uses AdaBoost algorithm. Speech convertor contains Single Point of Contact (SPOC) text to speech system [20] based on Hidden Markov models (HMM) for the speech recognition process. Speech convertor welcomes user's voice messages.

3. Information Collector Module (ICM)
ICM extracts personal information like gender, height, weight, waist, skin color tone, favorite color, flat footed, foot size, high sole, and event nature from the facts. ICM sends collected information (images and voice messages) to the item category selector module (ICSM). ICSM further handovers data to ICM.

4. Image Processing Module (IPM)
Based on different image processing techniques various features are extracted like skin tone, eye color, and face feature (lips, eyes) shape detection (see Fig.3). Image is converted into gray scale for building a histogram. Skin tone is checked to Von Luschan Chromatic scale [21]. Red Green Blue values of these types are determined and stored for future reference. Image is converted into HSV color space [22], [13]. The skin segments use predefined ranges of Hue, Cb, and Cr components of image. The ranges defined for these components are: Hue: 0.01 to 0.1, Cb: 140 to 195, and Cr: 140 to 165. The most frequently occurring Red Green and Blue components of skin region are determined for selecting the corresponding skin tone. Applying accessories, different facial points are extracted using image processing toolkit. Facial features are extracted based on Viola Jones's algorithm [23].

5. Items and Category Selector Module (ICSM)
ICSM manages different categories of wearable items according with the nature of events like causal (sport), or formal (wedding, party, and convocation), climate (cold, warm, etc.) size (small, medium, large, extra-large), height, gender, and body structure. Wearable items like dress, shoes, jewelry, bags, hairstyles, cosmetics, make up get ups and glasses are also selected. ICSM selects the wearable items category on the basis of event nature and personal information. ICSM accedes the knowledge base (KB) that retains record of wearable items according with color, size, and events. The queries are directed to the inference engine that consults KB and displays the list of inquiring items.

The inference engine consults inference rules for retrieving the commodities. Each rule is started with 'Startrule', describing the item category for which the rule is relevant. The end of the rule is indicated with 'Endrule'. For instance:

```
Startrule "Wedding dress suggestion"
  If user(Trial) = X
     user (X) = "customer"
     gender(X) = "female"
     selectetedItem (X) = Y
     requestedItem(Y) = "dress"
     skinTone (X) = fair
     height (X) = 42
     eventCategory (Y) = "wedding"
     color (Y) = deep red
     requestStuff (Y) =crinkle chiffon
     requestStyle(Y)= Lehenga
     priceRange =50000 /*Local currency */
  Then
   Display( Item(Y) style(Y) ← LehengaListsDeepRedcolor)
   Suggest(X) ←  Display (newlyFashionedList( dress, color)
Endrule
```

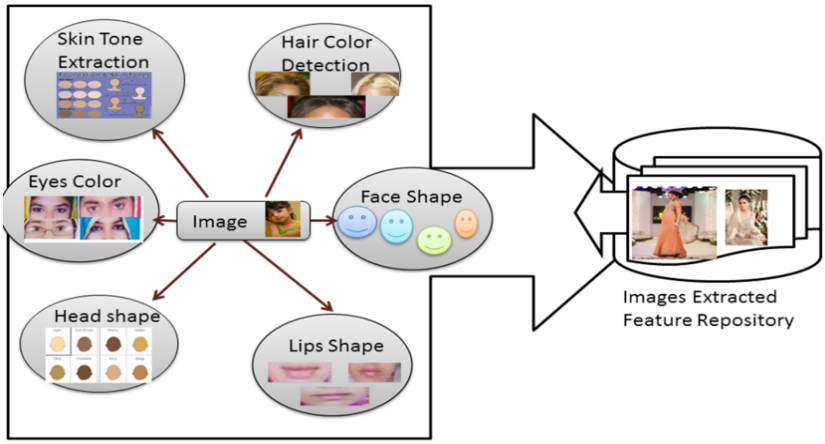

Fig. 3. Image Processing and Feature Extraction Module

The inference rule determines to display all related items that match with the required characteristics and in addition suggests others according with the user's features (see Fig. 4). Analogous rules are used for the suggestion of shoes, glasses, hairstyles commodities. There are other rules to suggest dresses for male with variety of different colors, price range event category: Party or Casual.

In addition, some questions can be enquired about the category of jewelry:

Earring: E category. A rule stores further facts about color, material shape of the earrings in the last price range. This rule is used to display rounded shape earrings with green stones. Similar rules are used to suggest for round shape face also.

Bracelets: B, Necklace: N, Rings: R, Watches: WW

Knowledge base is populated by storing similar rules for other commodities of female and male bags according to color, price range, category (HandClutch HC, HandBag HB ,Gents Wallets GW, . .).

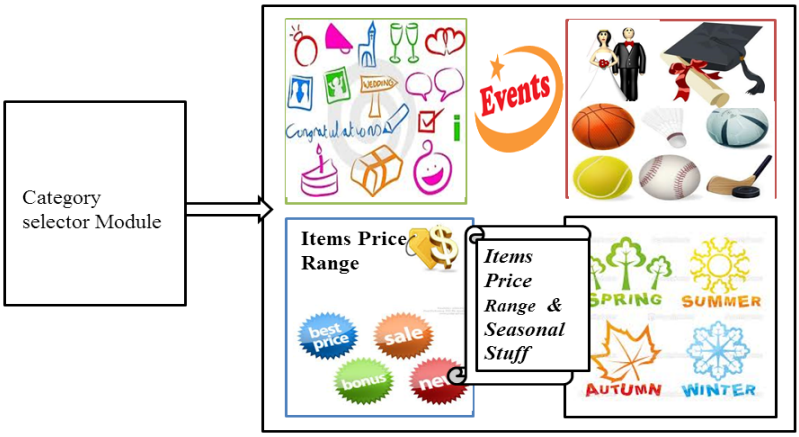

Fig. 4. Category Selector Module

6. Decision Module

DM receives a list of selected wearable items from KB. This list is built according to the precise information that ICMS provides to the knowledge base. When no precise match exists according to the user's requirements, the inference engine takes advantage of the reasoning ability delivers same item within same price range size but with different contrasts. After acquiring the image DM checks item by item on users snapshot forwards those snapshots to the decision dispatcher module (DDM).

7. Decision Dispatcher Module

Before displaying any result to the user, DDM has to perform two main tasks: **1.** DDM saves all the images that he received from the DCM into the local decision support system database. By means of the Hamming Distance Algorithm (HDA), DSS makes comparisons among different images, selects one of them that well-matched with the appearance of the user [17]. All images are sending to the convertor selector that converts the image description into text directs to PV. PV hover text information to the speech synthesizer that notifies user by describing verbal explanation about all images. **2**. DDM directs the selected items number bar code to the counter system, so that they can pack accessories for the user, prepare payment receipt for the user.

TASA-Blind is developed within Dot Net platform. For speech assistance, SAPI 5.4 and MS Speech DLL are used for the voice support TTS to STT conversion. AForge.NET DLL is used for capturing image.

Preliminary Evaluation. The speech grammar defined for the system works only with the expected words. Table 1 provides the standard user voice commands. Thus, the users know the vocabulary in order that they can properly interact with the system.

Let's consider a shopping mall, where Tania is a blind girl who wants to purchase a dress for a wedding event. When Tania enters into the trial room, TASA-Blind welcomes her, asking if she wants to interact with the Speech based interface: Please, select *Speech based* item section otherwise select *Manual selection*. She selects the former. System asks whether wants to visit *ladies* or *gents* items section, Tania says *ladies* section. Then camera captures her image to catch her skin tone, facial appearance, forwards this information to the system. After capturing image and height from the laser device, system navigates to the *ladies* items. Tania selects *Dress,* and then she is guided either to buy dress according to event nature or according to the seasonal stuff. The system moves to the next section that displays Event Based Selection.

Table 1. Speech Grammar Interface

Voice Commands	Actions and Description
Hello?	Welcome! For visiting ladies item section say : **Lady section** Otherwise say: **Gents section.**
Lady section	For capturing snapshot, please say : **Capture snapshot**, otherwise :**Skip snapshot**
Gents section	For capturing snapshot, please say : **Capture snapshot**, otherwise :**Skip snapshot**
Capture snapshot	Thank you, your picture has been already captured Now, Select desiring items, saying : **Dress, Shoes, Bags**, . . .
Skip snapshot	Ok! your snapshot will not be captured. You can select desiring items, saying : **Dress, Shoes, Bags, . . .**
Dress	When you want to choose dress according to events, say : **Event** , otherwise say : **Seasonal stuff**
Event	Now, select the event type, saying : **Wedding, Party, Sport . . .**
Wedding	For Mehndi dresses say : **Mehndi Wedding** In South Asia, Mehndi is a cultural social event before wedding day. For wedding day function say : **Wedding day**
Mehndi Wedding	Mehndi dresses are displayed and description item manufacturing material style is announced. Do you want to visit other items or **Close the session**?
Seasonal Stuff	If you are interested in season stuff, say it : **Summer, Winter, ,.**
Close	Gladly and cordially to receive you!

Now, the system asks her the kind of event. Tania Says: wedding. *As wedding is included within the set of events (wedding, sports, party, formal or casual event, . .*), the system asks her about any particular activity in wedding. Figure 5 shows the retrieved result of yellow Mehndi dresses from accessories stored in KB. All dresses taken out from KB are forwarded to DCA that acquires Tania's image from the user

data base, putting on all dresses to the Tania's image according to the acquired height that helps to suggest best size and fitting for Tania. Snapshots of Tania are transferred to the decision support system (DSS) of DDA that explains each dress appearance in terms of new trends like heavy embroidery with quelott, light ribbons work with trouser, stone work frock with pajamas by means of speech synthesizer. DSS also makes comparisons between all images, and informs Tania that the dress with light stone work is looking more beautiful than other dresses. Finally, Tania selects a slightly embroidered dress with long shirt trouser for a Social function. DDA order placement module sends the nominated dress bar code to the counter system places order for the dress. System inquires Tania if she would like to buy another item. Tania says : No so system thanks Tania for using TASA-Blind, and asks her to make payment at the counter. In this way, our system provides the impressive shopping assistance to its users.

Fig. 5. Yellow Mehndi Dress for Tania suggested by TASA-Blind

4 Conclusion and Future Work

TASA-Blind is an attempt for social integration. It helps blind to share ideas frankly, to shop individually, raise social activities and pay attention towards their look and personality. C Dot.Net platform is used for the development of TASA-Blind that supports both speech recognition and synthesis. Rules based on Predicate logic are used for the extraction of desiring wearable items from the base of rules. Additionally, a decision support system based on Hamming distance algorithm for making comparisons between user's suitable get ups. Speech synthesizer notifies users by describing verbal explanation about images, so user can easily make decision according to his/her own choice. As a future work, we will extend the vocabulary in order that visually impaired can talk freely instead using limited words.

References

1. Jacquet, C., Bellik, Y., Bourda, Y.: Electronic locomotion aids for the blind: Towards more assistive systems. Intelligent Paradigms for Assistive and Preventive Healthcare 19, 133–163 (2006)
2. Visual impairment and blindness-Fact Sheet (282), http://www.who.int/mediacentre/factsheets/fs282/en/ (visited on February 2012)
3. Blindness and Low Vision, Fact Sheet (2013), https://nfb.org/fact-sheet-blindness-and-low-vision (accessed March 19, 2013)
4. Slavík, P., Němec, V., Sporka, A.J.: Speech based user interface for users with special needs. In: Matoušek, V., Mautner, P., Pavelka, T. (eds.) TSD 2005. LNCS (LNAI), vol. 3658, pp. 45–55. Springer, Heidelberg (2005)
5. MarkkuTurunen, H., Soronen, S., Pakarinen, H.J., et al.: Accessible multimodal media center application for blind and partially sighted people. CIE 8(3), 16 (2010)
6. Alonso, F., Fuertes, J.L., Gonzalez, A.L., Martinez, L.: User-interface modelling for blind users 5105, 789–796 (2008)
7. Jian, Y., Jin, J.: An interactive interface between human and computer based on pattern and speech recognition. In: ICSAI, pp. 505–509. IEEE (2012)
8. Verma, P., Singh, R., Kumar Singh, A.: SICE: An enhanced framework for design and development of speech interfaces on client environment. IJCA 28(3), 1–8 (2011)
9. Bigham, J.P., Lau, T., Nichols, J.: Trailblazer: enabling blind users to blaze trails through the web, Sanibel Island, Florida, USA, pp. 177–186. ACM (February 2009)
10. Mahmoud, T.M.: A new fast skin color detection technique. World Academy of Science, Engineering and Technology 43, 501–505 (2008)
11. Sangho Yoon, M., Harville, H.: Baker, and N.Bhatii. Automatic skin pixel selection and skin color classification. In: Image Processing, pp. 941–944. IEEE (2006)
12. Siddiqi, M.H., Farooq, F., Lee, S.: A robust feature extraction method for human facial expressions recognition systems. In: IVC 2012, NZ, pp. 464–468. ACM (2012)
13. Oliveira, V.A., Conci, A.: Skin detection using hsv color space. In: Pedrini, H., Marques de Carvalho, J. (eds.) Workshops of Sibgrapi, pp. 1–2 (2009)
14. Shih, F.Y., Chuang, C.-F.: Automatic extraction of head and face boundaries and facial features. Information Science 158, 117–130 (2004)
15. Wechsler, H., Kidode, M.: A new edge detection technique and its implementation. Systems, Man and Cybernetics and Transaction on IEEE 7(12), 827–836 (1977)
16. Lee, J.S., Kuo, Y.M., Chung, P.C.: The adult image identification based on online sampling, pp. 2566–2571. IEEE (July 2006)
17. Aasim, K., Muhammad, A., Martinez-Enriquez, A.M.: Intelligent Implicit Interface for Wearable Items Suggestion. In: An, A., Lingras, P., Petty, S., Huang, R. (eds.) AMT 2010. LNCS (LNAI), vol. 6335, pp. 26–33. Springer, Heidelberg (2010)
18. Human Height Measuring, http://www.acuitylaser.com/products/category/human-height-measuring (visited on March 2013)
19. Fooprateepsiri, R., Kurutach, W., et al.: A fast and accurate face authentication method using hamming-trace transform combination. IETE Technical Review 27(5), 365 (2010)
20. BalaMurugan, M.T., Balaji, M., Venkataramani, B.: Spoc-based speechtotext conversion. National Institute of Technology, Trichy (2006)
21. Terrillon, J.-C., Akamatsu, S.: Comparative performance of different chrominance spaces for color segmentation and detectionof human faces in complex scene images. Vision Interface 99, 1821 (1999)
22. Patil, Y.M., Patil, M.M.: Robust skin colour detectionand tracking algorithm. IJERT 1(8), 1–6 (2012)
23. Paul, S.K., Uddin, M.S., Bouakaz, S.: Extractionof facial feature points using cumulative histogram. IJCSI 9 (2012)