Feed Forward Neural Network Training Based Interactive Shopping for Blind

Farzana Jabeen, Aslam Muhammad
Department of Computer Science
University of Engineering and Technology
Lahore, Pakistan
farzanajbn@yahoo.com, maslam@uet.edu.pk

Martinez Enriquez A.M
Department of Computer Science
CINVESTAV-IPN
D.F, Mexico
ammartin@cinvestav.mx

Abstract— **Today shopping markets pay attention towards customer needs and services. Unfortunately the blind and vision impaired person are still incapable to access these environments without reliance. Assistive technology is trying to sway the living style of the blind by introducing support systems for routinely actions like reading, writing, walking, Web surfing, and shopping. However, still the blind have to count on others for personal accessories shopping. To overcome this problem, we designed and developed a feed forward talking accessories selector. Our system is trained using feed forward techniques with a feature level block-based multi-focus image fusion method to provide suggestions regarding product selection, fitness, and color combination, for instance in dress and jewelry. The evaluation of our system takes into account specialists opinions, such that statistical analysis shows similarity between both.**

Keywords- shopping assistance, neural network, speech technology.

I. INTRODUCTION

According to the recent research, shopping is used as retail therapy to lessen the stress [1]. Window shopping is the common way that most of people adopt to get relax, entertained, or just discover new fashion and trend [2]. According to a survey, online shopper has increased from 172.13 million to 191.11 million in 2013 [3]. Technology has changed the shopping way and shopper styles too. But, the dependency of blinds on sighted persons for shopping is not changed yet. When sighted shopper go for buying, they search for items of interest, select a few for comparisons, and try them to check fitness and buy if any is the best for them after making consult with the dear one [4]. Whereas, the blind have to wait at info desk, so that any attendant is assigned to them [5]. So, they have to face the annoying and frustrated behavior of the assistant when they took more time in the product selection. No one can deny the problems that a blind has to face in his daily life. World Health Organization surveys show that the total estimated number of blind persons is 40-45 million [6] and only in Pakistan, the total estimated number of blind person was 1.25 million in 2003 and it may become 2.4 million till 2020 [7].

Today advancement in technology has changed the picture of blind person life. Using white cane, RFID based navigational sticks, screen readers, magnifier, and trail blazer, blinds can find their path easily and get access to reading, writing, and Web surfing. Supermarket shopping has become possible thanks to touch screen device, path dependent robot, and bar code scanner. But, still the blind hesitate to be social and attend formal functions. Due to lack of dressing sense and mismatch ability blind person used to stay at home instead of attending social events.

Like visionary person blind also need suggestion during shopping because whenever we consult with someone trust worthy it gives relief that we are buying the perfect product that suits us. Then, we should not forget the blind persons who even cannot see their appearance in mirror. It is true, the blind cannot understand product details in terms of color, stuff, shiny and dull material but they really want to buy perfect trendy and eye catching products.

To provide the optimal suggestion during shopping, we designed and developed a Talking Accessories Selector Assistant for blinds, which is a blend of speech and image processing, a knowledge base system, and artificial neural network (ANN). F2TASA-Blind provides speech communication to encourage the blind to express needs willingly and choose fashion items that are in well-matched to their personality. The aim is to make the blind self-reliant, confident, and dynamic in shopping of every commodity. Section 2 introduces the literature review. Section 3 illustrates the designed and implementation of our platform, as well as the validation of our proposal. Section 4 summarizes contributions and offers future research.

II. RELATED WORK

An image enhancement approach with the help of neural network classifier is used to provide mobility assistance for severe visual impaired persons in an urban environment [9]. Images are enhanced with high saturation color values. NN classifier is used to classify objects after analyzing the input feature vectors, labeling the identified objects according to their properties and assigns a color. *A hand gesture recognition system* is developed with fuzzy rule based approach and ANN to help deaf person [10]. Two layers HRCNN architecture is used to extract the rules from the network parameters. It recognizes the hand gestures that stores in the knowledge base for the NN training. Data base consists of 90 sign words from the Taiwanese sign

language. *An expert system* based on feed forward back propagation NN [11] comforts people by suggesting them suitable perfumes. After conducting the interviews with five best perfumes seller's authors identified 11 influential variables. Fuzzy Delphi method assigns the significance with these parameters, and then system is trained using ANN on the previous purchase data of 2886 person.

An intelligent tutoring system is developed to test the student learning ability with the help of neuro-fuzzy synergism [12]. The model consists of a fuzzifier, a fuzzy relational system, a fuzzy aggregation network, and a de-fuzzifier. The student model is implemented as continuous connections of these networks; each one provides fuzzy information about student's behavior. Fuzzy aggregation network is based on back propagation NN training. The student's knowledge level and cognitive abilities in a domain area is evaluated. *A generalized regression neural network* (GRNN) based system for the diagnosis of hepatitis B germs [13] has three layers for training, such that observes the behavior of Antigen (AgHBs), anti HVC, and anti VHD as input.

Back propagation ANN model is used along with genetic optimization for skin detection on human images [14], which successfully separates the pixel represented skin from others. *ANN based face recognition system* [15] is developed to provide security. The architecture consists of 960 inputs, 94 outputs, and 200 hidden layers. The trained system recognizes the face by removing noise.

All above systems assist the blind in different era of life. But when we browse for personal accessories shopping assistance, few systems are hardly found.

III. FEED FORWARD TALKING ACCESSORIES SELECTOR ASSISTANT FOR BLIND: F2TASA-BLIND

F2TASA-Blind system helps customers in choosing outlets without human assistant. To provide best suggestion, system is trained using feed forward propagation NN.

A. Design and Architecture of the System

In order to customers can easily and frankly share needs and requirements without human interruption support and to provide uninterrupted communication, a trial room is specially configured with physical components (see Fig.1):

- Web cam for taking user's snapshot, information that is stored into the image repository (IDB).
- Gait detector sensor determines the foot size and walking style of users [8], once user starts walking with bare footed, pressure sensors measure the different foot positions, providing size, flat foot, etc.
- Laser range finder device (Acuity AR 1000) is affixed in the trial room ceiling. The laser beam measures the height of beneath standing person.

- Speaker and microphone are part of HC Interface to provide precise verbal instructions to customer regarding functionality of the whole infrastructure.
- Work station coordinates the whole system and answers to the user's requirements.

The principal components of our platform are :

a. Active Insight Module (AIM)

AIM is liable to receive sensory information (image, data) and user's voice responses (see Fig. 2). AIM behaves like an interactive intermediate module with other component of F2TASA-Blind. The principal components are:

- Transient 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.
- English speech synthesizer transforms text into speech using synthetic human voice [16]. It recognizes speech lexicon by a predefined grammar.

b. Convertor Selector Module (CSM)

As input and output are temporarily reside in AIM, CSM receives acuity input from AIM that is being located in the I/O buffers. CSM performs several tasks : - saves a copy of user's snapshot into the IDB repository; - checks the input format either an image or speech; according to the input format, CSM directs data converting either text2speech or image2text; image2text conversion is based on Hamming Trace Transform (HTT) method [17].
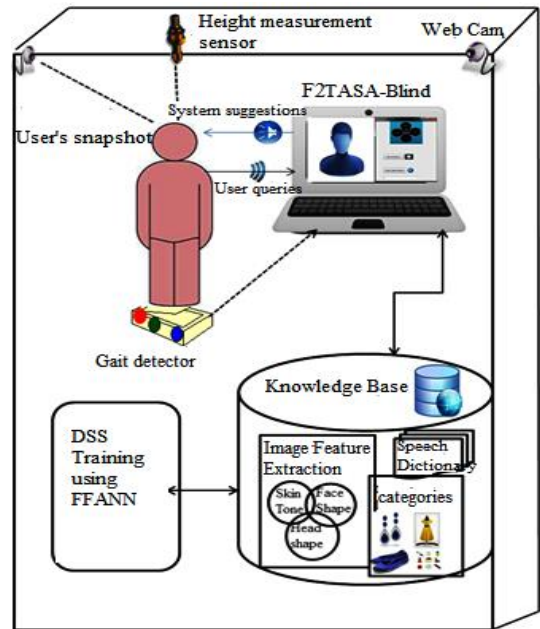
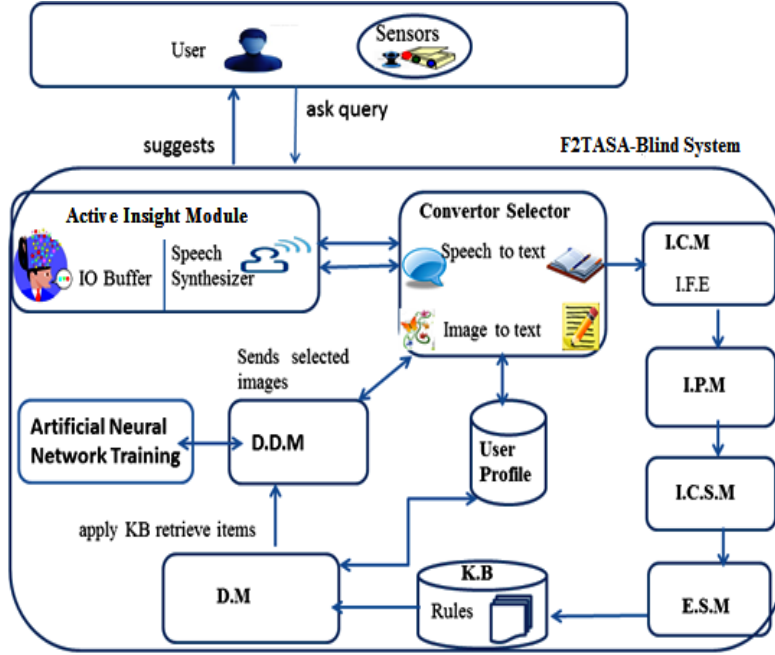

Fig 1: Model Schema of F2TASA-Blind

Fig.2: Architecture of F2TASA-Blind System

c. Information Collector Module (ICM)

ICM extracts user and product information. ICM sends collected information (images and voice messages) to the item category selector module (ICSM). ICSM further transfers data to ICM. Table 1 shows verbal information received by ICM.

Table 1. Verbal information

Parameters	Types
Gender	Male, Female
Age	(teen, adult, mature, old)
Favorite color	red, orange, green , black, purple, blue, white
Event type	wedding, party, educational
Product name	Dress, shoes, Jewelry, bags

d. Image Processing Module (IPM)

The user's items selection depends upon different physical features along with necessary verbal and sensory input as described earlier, e.g., skin tone unifies with dress colors; height with exact fitting and size; similarly face shape with the jewelry and glasses selection; type and tone of skin, and lips shape are major factor for the cosmetics suggestion. Table 2 shows the dependence between physical features and items selection. Thus, automatically features extraction plays a key role in the performance of F2TASA-Blind.

IPM detects the features from user's image: face shape, skin type and tone, eyes colors, lips shape, and neck height. First

image is converted into gray scale for building a histogram. Face shape is based on edge detection techniques to extract the contour [18].

Template matching technique [19] is the most effective and efficient for this purpose. After detecting the contour of the face, templates are used to extract the face shapes (round, oval, rectangular, and heart shaped) [20]. Skin type is defined as normal, oily, dry, and combination. To detect skin type, cheeks, nose, and chin region is compared. So skin type is extracted using tactile sensor [21]. Skin tone is checked to Von Luschan Chromatic scale [22] and calculated with the help of HSV color space [23]. Facial features like eyes, lips, hair color, and head shape are extracted using image processing toolkit, based on Viola Jones's algorithm [24] which is further divided into four stages: Haar features, images sub region, features selection, and circular shapes and colors.

Table 2. Items suggestion dependency upon physical features

Item Name	Auto extracted Features
Dress	Skin tone, Height
Jewelry	Face Shape, Height, Skin Tone
Hair Style	Face Shape , Head Shape
Cosmetic	Skin tone. Face Shape, Skin Type
Glasses	Face shape, Skin Tone, Distance between eyes and nose

e. Items and Category Selector Module (ICSM)

ICSM manages different categories of wearable items according with the nature of events like causal (sport), formal (wedding, party, and convocation), climate (cold, warm) size (small, medium, large, extra-large), height, gender, and body structure. Items like dress, shoes, jewelry, bags, cosmetics, and glasses are also selected. ICSM selects the items category on the basis of event nature and personal information (gender, age, height, skin tone). We implemented a knowledge based system (KBS) composed by: an Inference engine (IE) which satisfies Queries (Q) by consulting inference rules stored in the base of rules (BR) and displays the resulted items list. The base of facts (BF) retains record of items satisfying Q. Inference rules are written in first predicate logic (without function symbols). The known facts are stored in the base of facts (BF). The rule pattern is: **If** <condition> **Then** <action>. The inference rule starts with **Startrule**, follows by <description−identifier>, which identifies the rule and describes for which the rule is relevant (e.g., Necklace suggestion). The end of the rule is indicated with **Endrule**. Each rule is checked to be free from lexical, syntactical, and semantic errors. For instance:

```
Startrule "Necklace suggestion"
 If  (Trial) = X
      user (X) = customer
      gender(X) = lady
      faceShape(X) = round
      necklength(X)=short
      selectedItem(X) = Y
      categoryRequestedItem(Y) = necklace
      color(Y) = green
      material(Y) = stone
      style(Y)= round
      priceRange = 1000 /*Local currency */
Then
Display(Item(Y) style(Y)←
RounShape(CollarNecklaceList
(color (GreenStones)) )
Suggest(X)←Display(newlyFashionedList(collar
neckalce, shadesColor)
Endrule
```

The antecedent of the rule concerns with the requirements like customer X, located in the Trial room, ready for selecting items ("Necklace suggestion"). The customer is a woman who requests for a necklace (requestedItem(Y)= necklace). The face shape is (round) , and neck length is (short), her required color is (green) and the preferred material is *stone,* with round shape, within price range of P (P is a practical value in Local Currency). The rule actions describe/display related items that match with required characteristics, besides suggestions about others that are according with user's features.

f. Decision Module (DM)

DM receives a series of selected wearable items from KB that is retrieved by IE. The list is built according to precise information provided by ICMS to KB. When no match exists, IE takes advantage of the reasoning ability and delivers similar items within same price range size but with different contrasts. After acquiring the image DM checks item by item on users snapshot and forwards those snapshots to the decision dispatcher module (DDM).

g. Decision Dispatcher Module (DDM)

DDM provides the result to its user after performing the following tasks:

- saves all images into the local decision support system (DSS) database after receiving from DM. DSS is using feed forward NN with techniques like feature level block based multi focus image fusion techniques [25] with Eigen Patch method [4]. Figure 3 shows the working of DDS with the NN training. DSS gets the user images from the local database, dividing them into number of blocks. The size of the block is very important so that we can extract the desiring features from that block of the image [26]. Selected blocks from the current user images contain face part with suggested product. Five features including spatial frequency, contrast visibility, canny edges detection,

variance, and energy of gradient of the blocks are calculated. These blocks are then fed to ANN.

- Comparison is made on the bases of texture, colors, product shape, and fitness. After the comparison of block from each image, trained NN is used to fuse any pair of images. With the help of Eigen Patch method and Harris corner detection [27], texture comparison and fitness is checked of the final fused images. All images are sending to the converter selector that converts image description into text directs to AIM. AIM forwards text information to the speech synthesizer to notify user the verbal description about images, so that user can share own opinion because may be there is a color clash between user and DSS selected item.

- To place order when user is contented with the suggested commodity, DDM directs selected items number bar code to the counter system, so they can pack selected accessories and prepare the payment receipt.
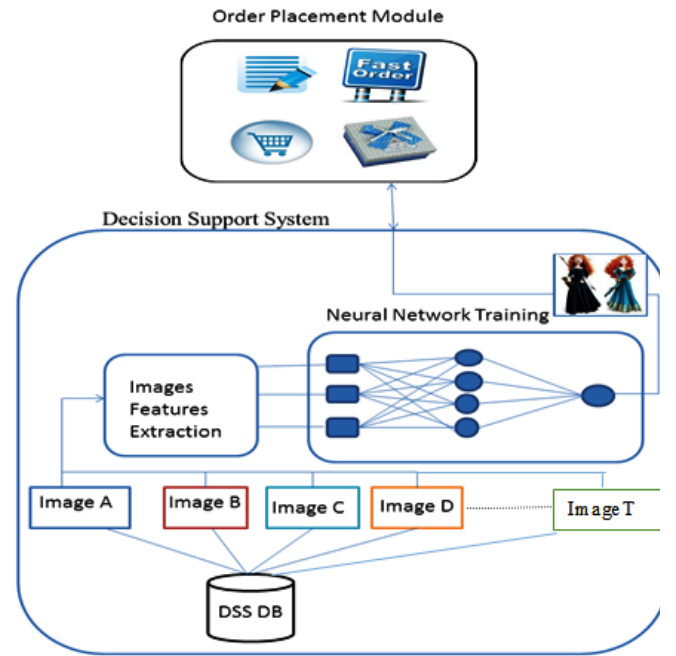

Fig. 3: Decision Dispatcher Module

IV. ARTIFICIAL NEURAL NETWORK TRAINING

In our platform, we are using multilayer feed forward neural network (FFNN) with feature level block based image fusion technique. The FFNN model is implemented with open source Neuron Dot Net c# engine. Three layers are designed: input, hidden, and the output layer. The input layer consists of 20 different images of a customer, with the same product with different colors, texture, and size. Hidden layers' neurons affect the efficiency and learning ability of FFNN, so it must be selected with great care. As no precise formula exists to determine the hidden layer [11], 2N+1, O

(N+1) provides a range to determine it, where N is the number of input, and O is the number of output [28]. In our case, N=20, O=5, so range is [41 to 105]. After some trials, 41 neurons are set as hidden layer. Finally hidden layer contains 41 neurons; output layer consists of 6 neurons. Learning termination preset iterations are 4850. The learning rate and threshold of mean square error remain constant as 0.01. From 20 images, it provides 5 fuse images based on texture, color, and size. Chi square test is performed on the resultant images to check the fitness. To check the correctness and accuracy of the NN model with image fusion techniques, statistical measures are computed [29], shown in Table 3.

Table 3: Statistical Computation

Images	Img1	Img2	Img3	Img4	Img5
MSE	0.00	0.01	0.05	0.00	0.01
MAE	0.028	0.035	0.016	0.149	0.124
R	0.985	0.965	0.997	0.957	0.988
SD	0.324	0.068	0.340	0.016	0.172
Percent result	98	99	100	100	95

To validate the NN training, a good approach is to analyze the behavior of MSE in each iteration. NN is trained enough when the value of MSE decrease with the increased number of epochs [28]. In our case, it starts from 0.27540 in the first iteration and ends at 0.01325 at the final iterations (see Fig. 4).
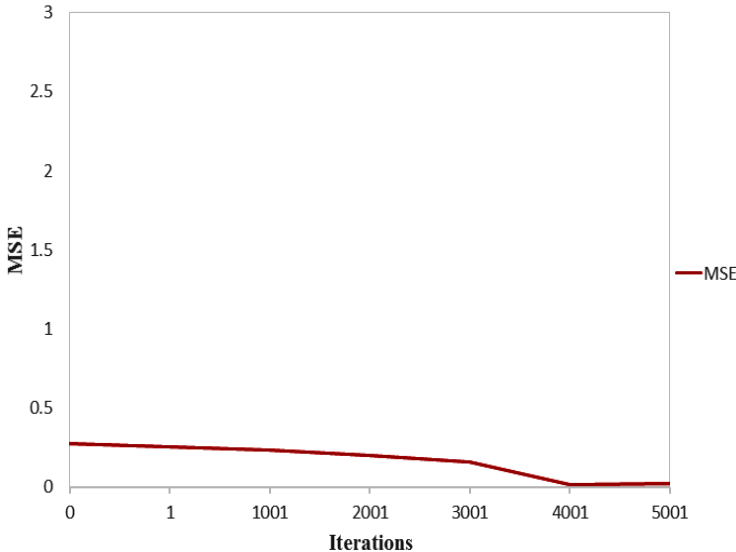

Fig. 4: MSE behavior during ANN training

The evaluation of the system is done with the opinion of five specialists contrasted with the selection makes by the system. After the interacting system, customers choose dress, jewelry, and cosmetics. A comparison is made with the system suggestion commodities with the selection

makes by the specialist, taken the coefficient of correlation. The average calculated value of the correlation analysis is 78.98 that show a good relevancy.

Case Study: Naima is a current trial room user. System prompts a voice message "describe your preferred gender items section" to guide her to start its selection. Meanwhile height measuring sensor measures the height and transfers info to the system. On Naima response, system moves to the ladies item section. AIM investigates about age. From her image captured, the physical features are extracted. After taking snapshot, F2TASA-Blind asks her to specify a category by saying dress, shoes, bag, jewelry, etc. Naima says "jewelry"; this response is recognized by the speech synthesizer and directs to the jewelry Section. "Please proceed by selecting jewelry category (earrings, necklace, bracelet, and rings)". Naima requests for a stoned necklace for a birthday party, price range of 2000 RS. After inquiring necessary info from Naima, AIM forwards it from the input buffer to CSM. CSM activates the desiring speech2text and image2text converter which transforms information from voice2text. User information (gender, age, height, item, size, color, price range) in text format is then sent to the Info Collector Module. Image processing module extracts physical features (skin tone, face shape, facial features) and forwards to the Item Category Selector module. ICSM directly communicates with KB and requests for all party wear necklaces with specified stone material. The fifteen retrieved images of stoned necklaces are then forwards to the Decision Dispatcher module which applies user snapshot to the retrieved results and forwards it to DDM. DDM saves images into its local repository, transfers them to DSS. DSS takes out these snapshots, divides them into a number of blocks, feed them as input to feed forward NN.

FFNN extracts features from each block and compared them to check the clarity, object shape, length and style. After training, five images are formed with image fusion technique. The five images are classified according to texture, style, and size of the necklace. Necklace suggestion mainly depends upon the face shape, height, age, event, and price range. Naima face shape is round, and height is small, so necklaces with choker lengths are the most suitable suggestion for her from DSS. These choker lengths images are forwards to AIM again. AIM speech synthesizer describes about the length, colors, stones shape, style, and price range of each necklace. System also informs Naima that choker length necklaces are suitable to her because of round face shape and small height. Finally, Naima selects a stylish choker length necklace with oval shape blue stone with price of 2000 RS (see Fig.5). System suggests Naima do not use perfume and any wet stuff on the necklace. The selected necklace item is sent to the order placement module which further forwards buying article bar code number to the counter for preparing receipt and payment. F2TASA-Blind thanks her for shopping and reminds to make payment at the counter.

Fig. 5: Selected Necklace for Naima

V.   CONCLUSION AND FUTURE WORK

F2TASA-Blind platform is blend of different technologies like speech and image processing, Knowledge base, Inference Engine, and ANN. Decision system supports on suggesting commodities to trend and user social functions. F2TASA-Blind is implemented in Dot net C#. Speech synthesizer has limitation on the accent, to identify acronyms, special texts. To overcome speech issues, our application uses two speech packages SAPI 5.4 Speech Library, MS Speech DLL for the TTS to STT conversion. Speech synthesizer communicates by verbal description about images, so customers make a decision in accordance with own choice. FFNN performance and accuracy is tested with the help of statistical measures. As a future work we will extend speech vocabulary and train FFNN with an efficient block level feature technique.

REFERNCES

1. Atalay, A. Selin, and Margaret G. Meloy. "Retail therapy: A strategic effort to improve mood."*Psychology & Marketing”* 28.6 (2011): 638-659.
2. Tauber, Edward M. "Why do people shop?". *The J. of Marketing* (1972): 46-49.
3. Number of digital shoppers in the United States from 2010 to 2018 (in millions) http://www.statista.com/statistics/183755/number-of-us-internet-shoppers-since-2009/
4. Zhang, Wei, T. Matsumoto, J. Liu, Maurice Chu, and Bo Begole. "An intelligent fitting room using multi-camera perception", *13th ICIUS*, pp. 60-69. ACM, 2008.
5. Lanigan, P. E., Aaron M. P., Andrew W. W., Dan R., and Priya "Trinetra: Assistive Technologies for Grocery Shopping for the Blind." In *ISWC*, pp. 147-148. 2006.
6. World Health Organization. Visual impairment and blindness, fact sheet n282. http://www.who.int/mediacentre/factsheets/fs282/en/.
7. Awan, Zahid Hussain, P. S. Mahar, and M. Saleh Memon. "Blindness and Poverty."*Pak J Ophthalmol* 27.3 (2011): 165-170.
8. Farzana J., Aslam M., A. M. Martinez-Enriquez, Z. S. Afraz, W. Talha."Speech Based Shopping Assistance for the Blind.", *6th MCPR 2014*, pp. 361-370. Springer LNCS.
9. Everingham, M. R., B. T. Thomas, T. Troscianko, and D. Easty. "Neural-network virtual reality mobility aid for the severely visually impaired." *the 2nd ECDVRAT*, pp. 183-192. 1998.
10. Su, Mu-Chun. "A fuzzy rule-based approach to spatio-temporal hand gesture recognition." *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on* 30, no. 2 (2000): 276-281.
11. Hanafizadeh, Payam, Ahad Z. Ravasan, and Hesam Ramazanpour Khaki. "An expert system for perfume selection using artificial neural network." *Expert Systems with Applications* 37, no. 12 (2010): 8879-8887.
12. Stathacopoulou, R., G. D. Magoulas, and M. Grigoriadou. "Neural network-based fuzzy modeling of the student in intelligent tutoring systems." *Neural Networks, 1999. IJCNN'99. International Joint Conference on*. Vol. 5. IEEE, 1999.
13. Mahesh, C., V. G. Suresh, and Manjula Babu. "Diagnosing Hepatitis B Using Artificial Neural Network Based Expert System." *infection* 3, no. 6 (2013).
14. Al-Mohair, Hani K., Junita Mohamad-Saleh, and Shahrel Azmin Suandi. "Human skin color detection: a review on neural network perspective."ICIC International, vol. 8, no. 12 (2012): 8115-8131.
15. Réda, Adjoudj, and Boukelif Aoued. "Artificial neural network-based face recognition." *Control, Communications and Signal Processing, 2004. First International Symposium on*. IEEE, 2004.
16. MT Bala Murugan, M Balaji, and B Venkataramani. Sopc-based speech- to-text conversion. National Institute of Technology, Trichy, 2006
17. Rerkchai Fooprateepsiri, Werasak Kurutach, et al. A fast and accurate face authentication method using hamming-trace transform combination. IETE Technical Review, 27(5):365, 2010.
18. Wang, Junyan, Guangda Su, and Xinggang Lin. "An effective approach to chin contour extraction." In *Advances in Intelligent Computing*, pp. 203-212. Springer Berlin Heidelberg, 2005.
19. Ping, Scott Tan Yeh, Chun Hui Weng, and Boonping Lau. "Face Detection Through Template Matching And Color Segmentation." *Nevim: Nevim* 89 (2003).
20. Kakumanu, P., Sokratis M., and Nikolaos Bourbakis. "A survey of skin-color modeling and detection methods". *Pattern recognition* 40.3 (2007): 1106-1122.
21. Lindahl, O. A., S. Omata, and K. A. Ängquist. "A tactile sensor for detection of physical properties of human skin in vivo."*Journal of medical engineering & technology* 22.4 (1998): 147-153.
22. Jean-Christophe Terrillon and Shigeru Akamatsu. Comparative performance of different chrominance spaces for color segmentation and detection of human faces in complex scene images. Vision Interface, 99:1821, 1999.
23. VA Oliveira and A Conci. Skin detection using hsv color space. In H.Pedrini, & J. Marques de Carvalho, Workshops of Sibgrapi, pages 1–2, 2009.

24. Sushil Kumar Paul, Mohammad Shorif Uddin, and Saida Bouakaz. Extraction of facial feature points using cumulative histogram. IJCSI, 9, 2012.
25. Pagidimarry, Madhavi, and K. Ashok Babu. "An All Approach for Multi-Focus Image Fusion using Neural Network." *Artificial Intelligent Systems and Machine Learning* 3.12 (2011): 732-739.
26. Mamatha, G., Shaik Abdul Rahim, and Cyril Prasanna Raj. "Feature-Level Multi-Focus Image Fusion Using Neural Network and Image Enhancement." *Global Journal of Computer Science and Technology* 12.10-F (2012).
27. Harris, C. and Stephens, M. A combined corner and edge detector. Proceedings of the 4th Alvey Vision Conference: pages 147—151. 1988.
28. Moraga, Claudio. "Design of neural networks." In *Knowledge-Based Intelligent Information and Engineering Systems*, pp. 26-33. Springer Berlin Heidelberg, 2007.