Chapter 2

Preliminaries and System Overview

2.1 Key Definitions

The definition of visual impairment has evolved over the years and varies across countries [96]. The literature contains multiple formal and informal definitions of visual impairment [97]. These are broadly categorised into disability-based functional definitions or measurement and quantification based definitions.

The World Health Organization (WHO) classifies visual impairment based on two factors: visual acuity and visual fields. Visual acuity is related to the clarity of vision. It is measured by comparing the vision of the person under test with the vision of a person with normal vision. If a person can see clearly at 20 feet what should normally be seen by a person with normal vision at that distance. Then, the person under test is said to have a 20/20 vision. Similarly, 20/70 vision means that the person is at a distance 20 feet and sees what a person with normal vision can

see from 70 feet. Visual acuity is measured using Snellen chart [98]. Snellen chart contains various alphabets of different sizes (refer Appendix A). They are viewed from a distance of 20 feet (6 metres). This chart tells us how well a person is able to see the letters and shapes from a distance of 20 feet. Each row of letters is given with a ratio which indicates the visual acuity needed to read it. The ratio corresponding to the smallest letter that a person can read refers to individual's visual acuity for that eye. The visual fields, i.e. the visual area that humans are able to perceive, while eyes are in a stationary position and looking straight at an object.

According to WHO and International Classification of Diseases-10 [97], vision is broadly categorised into normal vision, moderate vision impairment, severe vision impairment and blindness as explained below.

- Moderate visual impairment is a visual acuity between 20/70 and 20/200 with best-corrected vision, or a visual field of no more than 20 degrees.
- Severe vision impairment is a visual acuity of visual acuity between 20/200 and 20/400 with best-corrected vision or a visual field of no more than 10 degrees.
- Blindness is a visual acuity of 20/400 or worse with best-corrected vision or a visual field of no more than 10 degrees.

Moderate vision impairment and severe vision impairment collectively fall under the category of 'low vision'. Low vision along with blind represents all vision impairment. Recently, India has adopted a new definition of the blind in accordance with WHO. According to the National Programme for Control of Blindness (NPCB), if a person is unable to count fingers from a distance of three metres (earlier stipulation of six metres) would be considered as blind.

In this dissertation, the word visually impaired refers to low vision as well as blind.

2.2 Visually Impaired Users and HCI

In the modern digital world, computers are playing a key role in daily activities. Despite this visually impaired are still unable to use the computers effectively. The assistive technology can help them to interact with computers and be a backbone of independence for millions of people with disability. Assistive technologies enable them to participate as a contributing member of the society [99]. Thus, it is essential for people with disability to learn and adopt such techniques.

In the framework of this dissertation, technologies helping visually impaired individuals to interact with computers are studied, and hurdles faced by them are discussed. Numerous assistive technologies have been developed to reduce and bridge the gap between computer and a user with visual impairment [100]. Alternative keyboards, speech recognition, braille based input devices like e-brailler, electronic notetakers [101], etc. are available for providing input to computers. Similarly, speech assisted screen reading systems, enlarging display like video magnifiers (formerly known as closed-circuit televisions, CCTV), braille output printer, refreshable braille display and tactile output device (non-braille) are available as the output devices. Some of these are illustrated in Figure 2.1. The popularity of assistive technologies among

FIGURE 2.1: Some assistive devices available for visually impaired users (a) Perkins SMART Brailler® (b) Braille Notetakers (c) Enlarging display (d) Braille embosser.

people with disabilities have certainly increased over time. However, the adoption rate of these technologies is still low among the visually impaired users [102]. We have studied and investigated the reason behind the lower adoption rate. Some reasons for lower adoption rate are briefed below.

- **Adoption of technology**: It has been a major issue with blind and visually impaired users. One of the reasons for the low adoption rate is the pace at which technologies are evolving. According to [103] the development of new technologies over old technologies is so fast that visually impaired users find it difficult to maintain the same pace. Replacement of old technologies by new ones is usually too rapid. Thus, leaving assistive technologies to play catch-up. Another factor that discourages its use is the long time delay between the rise of new technology and the availability of assistive devices to the visually impaired users.

- **Cost factor**: According to facts presented by WHO, 90% of visually impaired people live in developing countries with low–and middle–income [104]. The average cost of living in a developing country is very less ($\approx 1/10^{th}$) as compared to a developed country. However, the cost of assistive devices for visually impaired users is quite high making it out of their reach [105, 106].

- **Lack of proper training and education**: Training is important regardless of technology. Despite some of the available solutions, visually impaired users are not a part of HCI due to lack of proper education and training. The study has suggested that a child's growth gets hindered if one moves towards technology without adequate knowledge and proper training [107]. Hence, blind and visually impaired users should be properly trained until they can use the chosen technology efficiently and effectively.

- **Socio-psychological factor**: A combination of various socio-psychological factors like social conditioning, control over surroundings, frustration, anxiousness, and social embarrassment discussed in [108] hinders the adoption of assistive technology. Blind and visually impaired users feel denial from society and some kind of humiliation and isolation.

  Most of the times visually impaired face issues due to their self-negligence. It is a state where person find that himself/herself cannot do anything and less of worth. Self-negligence has a negative impact on their mental thoughts which often leads to a feeling of loneliness and social isolation [109]. Due to such a feeling, they often lack/resist to try new method and technology.

- **Issue due to graphical user interface**: The introduction of graphical user interfaces (GUI) has presented a big obstacle for the blind and visually impaired users to effectively use computers [110]. GUI makes use of icons or other visual indicators to interact instead of using only text via the command line [111]. It allows users to directly manipulate the objects by performing actions like selection, dragging, etc. This action requires visual feedback (i.e. the location of the pointing device in space) making it difficult for the blind and visually impaired. The text part of the screen can be made available using Braille and speech output. However, the icons and buttons are not accessible to them [7]. An attempt was done in [112, 113] to make it accessible to visually impaired users using musical tones and synthetic speech. According to Boyd [114], blind and visually impaired users face three issues to access GUI/computer.

  i) Pixel barrier: The information of the screen is stored in a special buffer called as screen buffer or frame buffer. The frame buffer is a hardware device that is part of the graphics card. The contents of the screen are

stored as pixel map in the graphics memory. Text to speech converter cannot read the pixel map. Hence, it is difficult to provide voice output of the screen.

ii) Mouse issue: Blind and visually impaired users are unable to use the mouse due to lack of vision. They do not get feedback about the position of the mouse pointer. Hence, they face difficulty in using the mouse as a useful input device for accessing GUI.

iii) Graphics issue: In a GUI, the screen contents are provided as an image. The user recognises the pattern based on previous experience. The topography and topology provide an additional hint regarding the object on the screen. Textual explanations of graphics are long, vague and difficult to understand.

Apart from these factors, sometimes the impact of prior experience with similar technologies also affects the adoption rate.

2.3 Design Considerations

Let us discuss the design consideration that should be taken into account while developing special interfaces for blind and visually impaired. Research has shown that the functional disorder of a particular sense enhances the other senses. Brain plasticity theory states that blind people have enhanced capability in the remaining modalities [115]. They have better hearing, smell and tactile capabilities [116]. It is often said that they develop this enhanced capabilities by utilizing the unused part of the brain. The brain of blind and visually impaired people are wired to enhance

other senses [117]. Hence, these enhanced capabilities should be utilized to provide access to computers.

Haptics is a form of interaction involving touch. It is the science of applying tactile sensation and control to interaction with computer applications. Users can receive feedback from computer applications in the form of felt sensations in the hand or other parts of the body. Different issues and suggestions related to the use of haptic technology for assisting visually impaired users are presented in [118]. The use of haptic as an alternate to vision is comparatively harder, and slower but allows visually impaired user a method of getting output from the computer.

Voice interface provides an alternate method of getting output from the computer. It is a technology that is already in use for such users. It has the capability to be used as an alternative to vision. A human eye can observe variations of light hue, brightness and contrast. Similarly, the human ear can understand a variety of sounds. Next, the human brain is capable of relating these sounds pattern with events, object, action etc. Voice interface has one of the greatest benefits that they do not require visual attention hence suitable for visually impaired users.

As discussed earlier, the loss of a particular sense is often compensated by providing the required information through the available senses [119]. The vision loss in blind and visually impaired can be compensated by combining haptic and auditory feedback. However, sometimes providing this extra information through the alternate senses may result in sensory overload. Hence, care must be taken to prevent sensory overload.

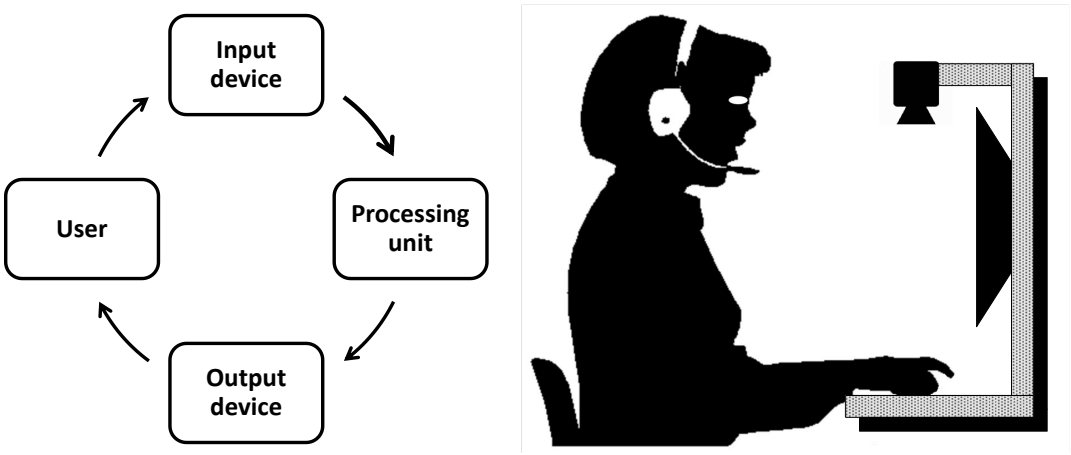

FIGURE 2.2: Abstract representation of the proposed interactive system.

2.4 Proposed Interactive System Framework

This section details the framework of the proposed interactive system. It consists of three interconnected research: user interface design, dactylology proposal and a recognition module.

2.4.1 User Interface Design

In conventional gesture-based interaction, a camera is installed in front of the user and user needs to pose gesture in space within its field of view. Blind people find it difficult to pose a gesture in space without support. Non-availability of feedback makes it even more difficult for them to orient and pose their gestures properly. Additionally, they also find it hard to hold the gesture perpendicular to the camera axis. This causes perspective errors and results in projective distortion.

As suggested in Section 2.3, haptic and voice interface should be used to overcome the loss of vision. So, we have provided a tabletop arrangement for posing gestures to make use of haptic/touch as alternative sense to vision, as shown in Figure 2.2. This arrangement not only provides haptic feedback but also provides support to the arms. Additionally, they help to maintain the gesture pose perpendicular to the camera axis. The set-up consists of a camera and it has been installed at the top to capture gesture posed in its field of view. Upon capturing a gesture, it is transferred to the processing unit which recognizes the same and decodes the command given by the user. Audio feedback is also provided to the user through headphone/speaker to ensure correct and reliable data entry.

Ergonomic aspects are also considered while designing this interactive interface. The form factor of the interface is also taken into account for better accessibility. The height of the table is made adjustable such that the shoulder is in a relaxed position

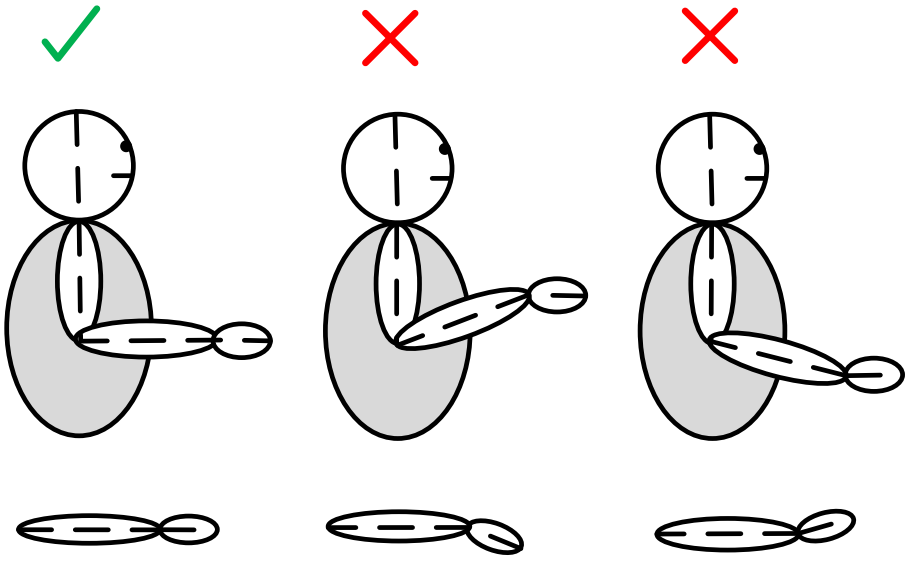

FIGURE 2.3: Illustration showing correct hand posing based on ergonomic aspects.

with wrist and forearm be in-line with it. Upward–and side–bending of the wrist should be avoided for stress-less postures.

2.4.2 Proposed Dactylology

Dactylology is defined as a technique of communicating/interacting with computers by symbols (also known as pose/gestures) made with the fingers. Symbols are made by varying hand fingers configurations to represent a letter, number and additional symbols.

In this dissertation, we have proposed a novel dactylology for blind and visually impaired people. The proposed dactylology is an outcome of user evaluation study performed with 25 blind and visually impaired users. In this user evaluation study, we conducted a quantitative rating analysis to choose optimal gestures. This rating analysis indirectly gives an indication of complexity levels of gestures in the set and their suitability for the users. More than 12,400 questions were asked and analysed in this quantitative rating analysis. Gestures in the dactylology are selected based on performance and preference measure metrics. Performance measure includes rating of gestures on the basis of easiness (C_1), naturalness (C_2), ease of learning (C_3), and reproducibility (C_4). A Likert scale (1=strongly disagree to 5=strongly agree) is used to rate gestures on questions related to four criteria. In preference measure, a popularity index is calculated to explicitly consider the popularity and preference of a gesture among blind users. Finally, dactylology is proposed using optimal gestures obtained from performance and preference measure metrics. Detailed discussion on the user evaluation study and the proposed dactylology is presented in Chapter 3. When a symbol is posed as per the proposed dactylology, the input device (i.e. camera) captures it and the captured frame is further sent to recognition module

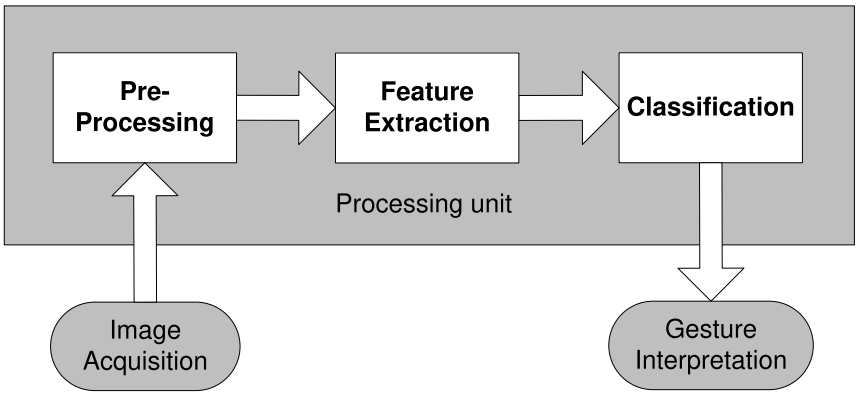

FIGURE 2.4: Abstract level representation of recognition module.

(i.e. processing unit) for recognition and interpretation. Discussion on recognition module is presented in the next section.

2.4.3 Recognition Module

Recognition module is one of the most crucial blocks in any vision-based interactive system. An outline of the recognition module is shown in Figure 2.4. It has three major blocks namely image acquisition, processing unit, and gesture interpretation. Image acquisition is the first block where optical flow method is used to obtain displacement vector between two consecutive frames. Based on this, frames which seem to be static are captured and processed further. The processing unit extracts distinctive features and performs classification using a rule-based classifier. Finally, the gesture interpretation block recognizes the symbol. The processing unit comprises of three stages—pre-processing, feature extraction, and classification.

2.4.3.1 Pre-Processing

Pre-processing is a technique in which an outcome is more appropriate for feature extraction than the acquired image. In this case, it is aimed to extract the hand from the input image acquired using the camera. Steps of pre-processing are illustrated in Figure 2.5. Details about these steps are discussed below.

1. **Illumination compensation:** Illumination variation is a challenging problem in a real-time vision-based system. It changes the appearance of an object significantly. In some cases, the difference induced by illumination is so huge that it makes skin segmentation task more difficult. The effect of illumination can be reduced by applying the Gray World algorithm [49]. This algorithm assumes that the average colour of the surface of the image is always achromatic. It means that the average colour reflected from an image surface is the colour of the illumination source. This condition may or may not be satisfied. Let an image $I(i, j)$ have size $M \times N$, where i and j denote the indices of the pixel location. Furthermore, let $I_r(i, j)$, $I_g(i, j)$ and $I_b(i, j)$ be the red, green and blue channels of the image, respectively. Then the average for each channel is

$$R_{avg} = \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} I_r(i, j) \tag{2.1}$$

$$G_{avg} = \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} I_g(i, j) \tag{2.2}$$

$$B_{avg} = \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} I_b(i, j) \tag{2.3}$$

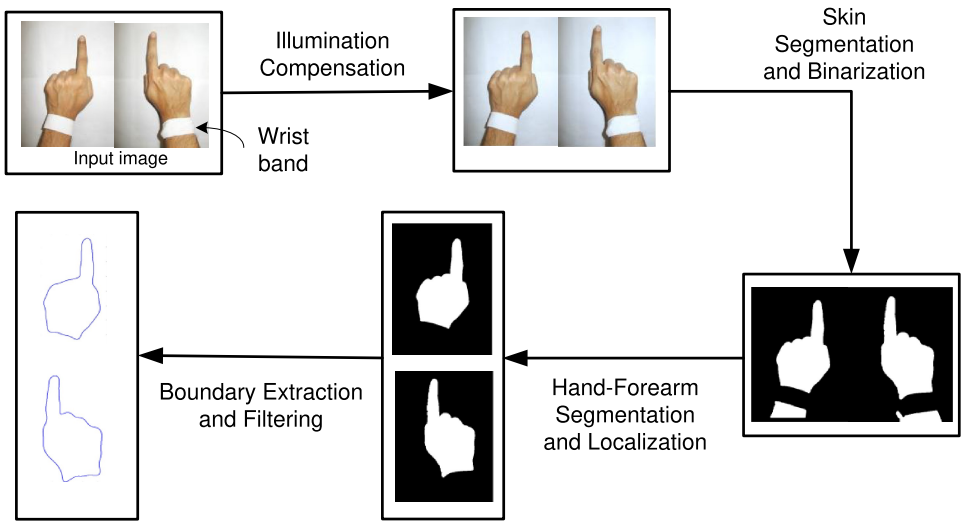

FIGURE 2.5: Pre-processing steps

where R_{avg}, G_{avg} and B_{avg} is the average value of R, G and B channels, respectively. Illumination estimate is obtained by calculating the mean of each channel average. The corresponding scale factor for each channel is calculated by

$$S_r = \frac{\text{Illumination estimate}}{R_{avg}} \tag{2.4}$$

$$S_g = \frac{\text{Illumination estimate}}{G_{avg}} \tag{2.5}$$

$$S_b = \frac{\text{Illumination estimate}}{B_{avg}} \tag{2.6}$$

where S_r, S_g and S_b is the scale factor for R, G and B channels, respectively. The compensated colour information for each channel (R', G' and B') of the input image is obtained by multiplying the original colour information with its corresponding scale factor Sr, Sg and Sb, respectively.

2. **Skin segmentation and binarization:** This step converts illumination compensated RGB image into binary hand mask (M_h). The basic idea of skin segmentation method is that the color of skin is unique and it forms a separate cluster in different color space. The boundary of these cluster is obtained empirically by analysing skin and non-skin pixels. Some of the examples are RGB [55], HSV [56], YCbCr [57], etc. It is found that skin cluster is compact as well as non-overlapping in YCbCr than other color spaces [57]. Hence, we have transformed the illumination compensated RGB image into YCbCr color space. Next, thresholding is performed using optimal range reported in [57]. These boundary based techniques work well in the constrained environment like uniform background, good lighting condition etc. However, in the case of unconstrained environment, skin segmentation becomes quite challenging. Variation of illumination, inter-personal differences and a wide variety of races, ageing, etc. make skin-color based segmentation even more challenging task. Apart from this method, we have also implemented some of the state of the art methods and those are briefed below.

    - Bayesian skin modeling: In this work, a Bayesian classifier is used to model the input image into skin probability map (SPM). Here, a histogram of skin and non-skin pixels are analyzed. Based on this histogram, the probability of a given pixel value (v) as the skin is obtained using Bayes rule [64] given below

    $$P(C_s|v) = \frac{P(v|C_s)P(C_s)}{P(v|C_s)P(C_s) + P(v|C_{ns})P(C_{ns})} \tag{2.7}$$

    where a priori probability ($P(C_s)$) and ($P(C_{ns})$) is assumed to be half.

    - Fast Propagation-based Skin Detection (FPSD): In this work [65], a spatial analysis of the skin probability map (SPM) is proposed to improve

the skin segmentation results which are obtained using conventional pixel-wise detection. Here, a distance transform is used for skinness propagation in a combined domain of luminance, hue and skin probability.

- Discriminative Skin Presence Features (DSPF): In this work, the texture-based discriminative skin presence features [66] are proposed to improve the skin classification accuracy. These textural features are obtained from the probability maps [64] rather than luminance channel.
- Multi-Seed Propagation in Multi-layer Graph (MPMG): In this work [67], an initial skin probability map (iSPM) is generated using Bayesian modeling [64]. Here, an image is represented as a multi-layer graph: image layer and cluster layer. Initial seeds are selected based on the generated iSPM (skin: high iSPM, non-skin: boundary and low iSPM). These seeds are propagated through the multi-layer graph using semi-supervised learning. This step is followed by a pixel-wise refinement for suppressing the non-skin pixels in the final SPM.

Further details and literature about the skin segmentation and binarization are discussed in Chapter 5. The impact of skin segmentation on the subsequent steps is also discussed in Chapter 5.

3. **Hand-forearm segmentation:** Hand-forearm segmentation is another important step of the recognition module. It is essential to segment and discard the forearm region from the extracted image because there is no distinct information. Additionally, non-removal of the forearm region often leads to variation in extracted features affecting the correct recognition.

   Most of the time hand-forearm segmentation is performed by imposing constraints on the users. Users are supposed to wear a white band which separate hand and forearm region. The hand is localised by assuming that hand is the

largest object in the frame grabbed. However, imposing restrictions on the users is not recommendable as it hinders the natural interaction mode. Hence, we present an automatic approach for robust hand-forearm segmentation using geometric features. In this work, circular and elliptical shapes are used to approximate the hand palm. Next, a wrist point detection method is proposed which is inspired and based on the observation of human hand anatomy. *Further details about the automatic hand-forearm segmentation methods are discussed in Chapter 5*

4. **Boundary extraction and filtering:** Boundary extraction and filtering step provide smooth and filtered boundary of the localized hand. The boundary of the segmented binary image M_h is obtained using a dilation operation. For this purpose, segmented binary image M_h is dilated by a structuring element S. Next, the set-differences between the dilated version of image and the original segmented image is calculated as below

$$\beta(M_h) = \{(M_h \oplus S) - M_h\}. \tag{2.8}$$

Here 'S' is a disk-shaped structuring element whose radius is equal to 3 pixels and '$-$' is the difference operation on the two sets. Further, the extracted boundary is filtered using a Fourier filter [33] so that the effect of boundary distortion and noise is reduced. Suppose $F(u)$ is the Fourier transform of the extracted boundary $\beta(M_h)$ which is represented by sequence of complex numbers $s(i) = x(i) + jy(i)$ and the low pass filter transfer function is $H(u)$; output is given by multiplication in the frequency domain, i.e. $F(u).H(u)$, where $H(u)$ is given by

$$H(u) = \begin{cases} 1, & \text{if } D(u) \leq A_0 \\ 0, & \text{otherwise} \end{cases} \tag{2.9}$$

where u denotes spatial frequency, A_0 is the cutoff frequency and $D(u)$ is the distance between the frequency origin and spatial frequency u. The value of A_0 is selected on the basis of repeated experiment [120]. It is observed that good approximation to the hand shape can be obtained with a relatively small value of A_0. In our experiments, we chose $A_0 = 40$ as this value of A_0 provided better approximation of shape and smooth boundary.

2.4.3.2 Feature Extraction

Feature extraction and its matching is the most crucial step in any vision-based gesture recognition. Feature extraction is a method of transforming raw/pre-processed input image into compact, informative and non-redundant features. These features are selected based on the distinctive properties of input image that help in differentiating between the categories of input image. In a hand gesture recognition, extracted features are usually classified into two categories—low-level and high-level

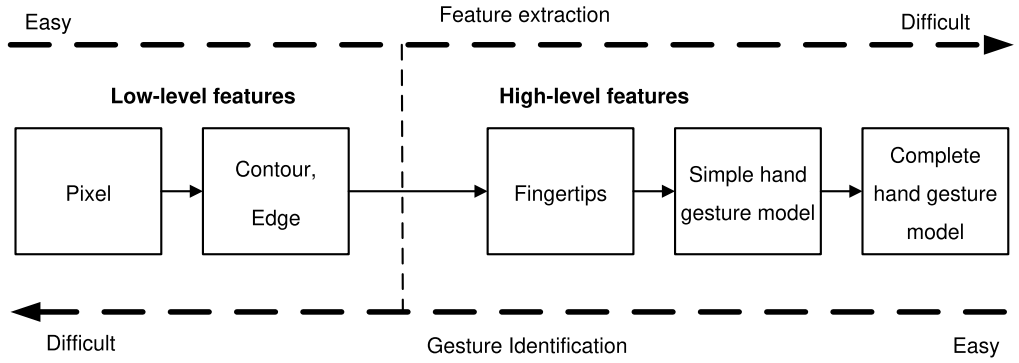

FIGURE 2.6: Features used in a hand gesture recognition system [121]

[121]. Low-level features include pixel, edge, edge orientation, histogram of oriented gradients (HOG), contour/silhouette features etc. These features are easy to calculate. However, in practical environment, even a slight variation in hand pose configuration causes variations in low-level features. With scale, translation and rotation variations, it is hard to recognise the gesture using only low level of features. High-level features include finger-tip attributes (e.g. position, location, and orientation), features related to structure of hand shape etc. These features are difficult to extract, but their recognition accuracy is quite good. High-level features are usually built on the top of low-level features.

In this work, we have proposed a new shape descriptor—reduced shape signature—for hand gesture classification. *Further details about the feature extraction methods are provided in Chapter 4.*

2.4.3.3 Classification

In this work, we have used rule-based classifier which classifies symbols according to the set of rules. Here, each classification rule is of form

$$r : (Condition1)\ \&\ (Condition2) \rightarrow y. \tag{2.10}$$

LHS of the rule is called the rule antecedent or precondition. Generally, it is a conjunction of several attribute tests. RHS of the rule is called the rule consequent and is also called as the class label.

The input test features are compared with the manually pre-coded rules. If the features match a particular rule, the corresponding gesture will be predicted at the

output. The classifier comprises of all such rules and hence, it turns into a simple look-up table.

Further details about the classification are provided in Chapter 4.

2.5 Concluding Remarks

In this Chapter, we presented some important definitions, preliminary study and overview of the proposed interactive system. In the preliminary study, we explored the hurdles faced by the visually impaired users in HCI. The factors like adoption of technology, cost factor, lack of proper training, and socio-psychology factors etc. are responsible for lower adoption rate of assistive technologies and hence one should overcome such hurdles. Some design considerations which are necessary from their perspective for the development of special interfaces are presented. Use of the alternate senses was suggested to overcome the visual impairment. But, one must take care of the sensory overload.

Further, we discussed the framework of the proposed interactive system. In this chapter, issues faced by visually impaired users while working with a conventional gesture-based interaction are discussed and a table-top arrangement is proposed as one of the solutions to it. The proposed arrangement make use haptic feedback as well as maintain the hand pose perpendicular to the camera axis. Next, we discussed the user interface design which is followed by a brief introduction to the proposed dactylology. Finally, a complete overview of the recognition module is presented and their details are discussed.