

Research Article

Video Analysis of Human Gait and Posture to Determine Neurological Disorders

Howard Lee,¹ Ling Guan,² and Ivan Lee³

¹ C Management Services Pty Ltd, CQU Melbourne International Campus, Melbourne, Australia

² Department of Electrical and Computer Engineering, Ryerson University, Toronto, Ontario, Canada M5B 2K3

³ Department of Computer and Information Science, University of South Australia, South Australia, Australia

Correspondence should be addressed to Howard Lee, hlee@ee.usyd.edu.au

Received 15 January 2007; Accepted 7 March 2008

Recommended by Alice Caplier

This paper investigates the application of digital image processing techniques to the detection of neurological disorder. Visual information extracted from the postures and movements of a human gait cycle can be used by an experienced neurologist to determine the mental health of the person. However, the current visual assessment of diagnosing neurological disorder is based very much on subjective observation, and hence the accuracy of diagnosis heavily relies on experience. Other diagnostic techniques employed involve the use of imaging systems which can only be operated under highly constructed environment. A prototype has been developed in this work that is able to capture the subject's gait on video in a relatively simple setup, and from which to process the selected frames of the gait in a computer. Based on the static visual features such as swing distances and joint angles of human limbs, the system identifies patients with Parkinsonism from the test subjects. To our knowledge, it is the first time swing distances are utilized and identified as an effective means for characterizing human gait. The experimental results have shown a promising potential in medical application to assist the clinicians in diagnosing Parkinsonism.

Copyright © 2008 Howard Lee et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

Parkinsonism or Parkinson syndrome is a clinical entity produced by several different etiological agents, and it is associated with a variety of pathological processes which damage the extrapyramidal system. The diagnosis is usually not difficult when the full clinical picture—tremor, rigidity, postural instability, and a decrease in spontaneous movement—is present. However, in the early stage, often show fragments of the total syndrome are evident in most of the patients and diagnosis then may not be completely certain [1–4]. Generally, follow up with the development of other symptoms and signs makes diagnosis possible.

Patients with Parkinson syndrome stand in a posture of general flexion with the spine bent forward, the head bowed, the arms moderately flexed at the elbows, and the hips and knees mildly flexed. They stand immobile and rigid, with a paucity of automatic movements and a mask-like face. Although the arms are held immobile, there may be a slow frequency tremor that involves the fingers and wrists.

This is often accentuated or even brought out once walking commences.

When walking commences, there is a restricted rotation of the trunk. As a result of the body being carried on the toes, the trunk bends even further forward, which pushes the center of gravity ahead of the foot support. This results in a propulsive gait with an inability to halt forward progression and risk of falling. One or both arms may fail to swing. The legs remain bent at the hips, knees, and ankles with reduced angular excursion at all the joints. The step cycle is lengthened in duration mainly due to an increase in the stance phase whereas the swing phase is reduced. The feet scrape and shuffle along the floor due to a reduced step height and a reduced stride length.

There is a disturbance in postural reactions due to abnormalities of central reflexes involved with postural adjustments. If the patient is pushed backwards or forwards, he may not be able to compensate with flexion or extension movements of the trunk, and he may fall precipitously. He often has trouble initiating gait after standing still or sitting

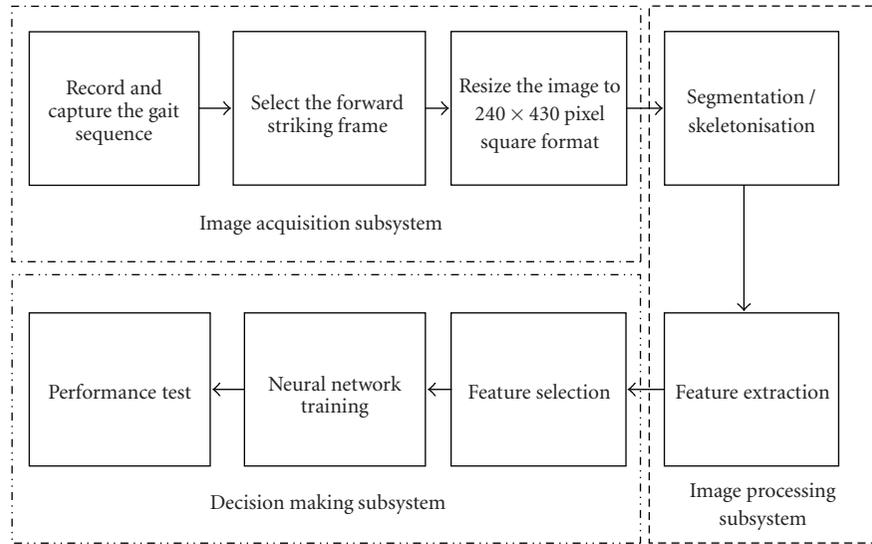


FIGURE 1: System diagram.

in a chair. The gait may become arrested by minor visual or proprioceptive stimuli and, if psychological stress is added, the patient may become “frozen”. On the other hand, it is well recognised that, in some patients, walking may be facilitated by different external stimuli. Trivial signals, for instance an object placed on the floor, may help initiate walking [4–6].

Parkinsonism patients have shown significant movement restrictions on the limbs which have not been adequately utilized in traditional diagnosis that it solely depends on the experience and judgment of the clinicians, which could be subjective and inconsistent. To provide accurate and quantitative measurements, two kinds of image processing diagnostic systems were proposed in the literature. One requires complex laboratory settings and body attachments such as motion marker systems and ground reaction force plate to measure various features for Parkinson syndrome, which may affect the gait movement [7–11]. The other is based on automatic analysis and recognition of human behavior by gait in videos recorded in an environment with relatively simple settings [12–17]. The second approach is of particular interest to this work.

Automatic analysis and recognition of human behavior by gait are subject to increasing interest and they have the unique capability to recognize people at a distance when other biometrics is obscured. Its interest is reinforced by the longstanding computer vision interest in automated noninvasive analysis of human motion. Its recognition capability is supported by studies in other domains such as medicine (biomechanics), mathematics, and psychology which continue to suggest that gait is unique. Current approaches confirm the early results that suggested gait could be used for identification, and now on much larger databases. Gait has benefited from the developments in other biometrics and has led to new insight particularly in view of covariates. As such, gait is an interesting research area, with contributions not only to the field of biometrics but also to the stock of new techniques for the extraction and

description of objects moving within image sequences. In biomedical applications such as analyzing neurological disorder and monitoring rehabilitation process after orthopedic surgeries, this could greatly reduce the hardware setup and give the patients the maximum possible comfort during diagnosis process. Recent survey papers on gait analysis gave comprehensive treatment of the subject, especially in the framework of two prominent gait analysis programs: Human ID at a Distance [18, 19], and Human ID Gait Challenge [20].

The work reported in this paper falls in the second category: automatic analysis and recognition of human behavior by gait. Based on the forward striking instant of the gait posture, the system measures, in a simple setting, features such as the swing distances of arms and legs as well as the angles of various joints to give an objective measurement to assist clinicians in determining potential Parkinsonism in the patients.

2. SYSTEM STRUCTURE

As shown in Figure 1, our video analysis system comprises of three subsystems: (1) image acquisition, (2) image processing and analysis, and (3) decision-making.

- (i) The main purpose of the image acquisition subsystem is to digitize and store the images from video sequences, so that the images can be processed and analyzed in the next stage.
- (ii) The image processing subsystem aims at extracting important visual information from the digitized images. Features such as swing distances and joint angles are extracted in this stage. The most effective features are then selected by a combination process of sequential backward selection (SBS)/general regression neural network (GRNN), and histogram analysis.

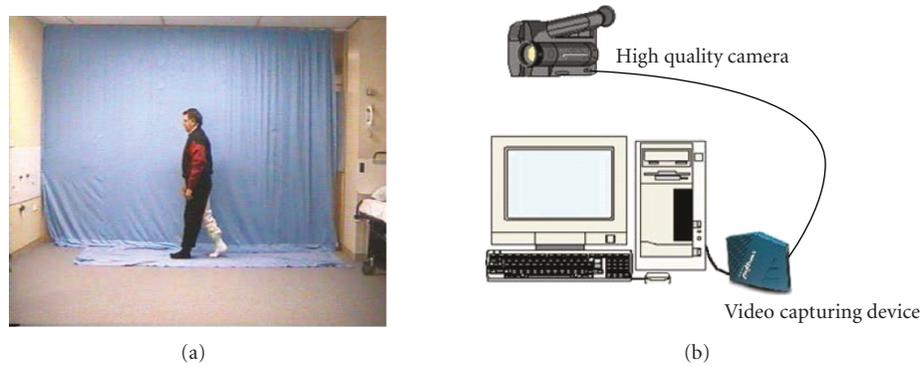


FIGURE 2: Image acquisition subsystem. (a) Example of the captured image (b) hardware settings.

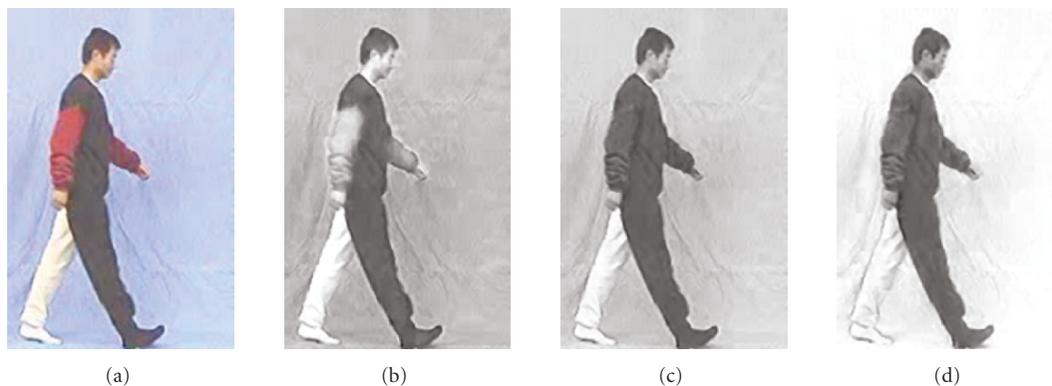


FIGURE 3: RGB components: (a) original image, (b) red component, (c) green component, (d) blue component.

- (iii) The decision-making subsystem is based on a feed-forward neural network, which is trained by the features selected previously. The network is then used to diagnose new neurological data.

3. THE IMAGE ACQUISITION SUBSYSTEM

In this work, the image acquisition subsystem comprises of two major steps: (1) video recording (2) image capturing and digitization. These two steps can be implemented independently or coexist at the same time.

Several distinctive colors are used on the tracksuit to highlight different parts of the body. Black color is used for the torso and the leg facing the camera. White color is used for the leg further away from the camera, and red color for the arms (see Figure 2). Rather than using expensive materials for the costume, the new suit was made out of tracksuit materials which are close to the normal outfit people wearing daily, and subjects tend to behave more naturally in the taped video. As a result, a more realistic gait patterns were observed, and more accurate data had been collected. In addition, this practical laboratory setting would be more beneficial in the context of medical research.

A high-quality portable S-VHS video camera is used to capture the walking sequence of the human subjects. In this

work, only the side-on views of the subjects are investigated. An example of the captured image is shown in Figure 2(a). The entire gait sequence was transferred to a computer using Videonics Python, a video capturing device as illustrated in Figure 2(b), and the relevant frames can be precisely selected for the experiment.

The special designed tracksuit and uniform colored backdrop were used to simplify the segmentation process. In the realistic system, these constrains should be replaced by the subjects wearing normal outfit and the experiment be carried out at any indoor environment. This may involve shooting in stereograph picture to obtain the distance information of the limbs from the camera, hence to perform spatial segmentation.

4. THE IMAGE PROCESSING SUBSYSTEM

Segmentation subdivides an image into its constituent parts or objects. The levels, to which this subdivision goes on, are depending on how much detail is desired for the overall function to be successfully completed. Before an image was segmented, it was cropped to the desired size, and then filtered by a low-pass filter to remove random noise and a median filtered to eliminate speckle noise. The resulting image is served as the input to the segmentation phase.

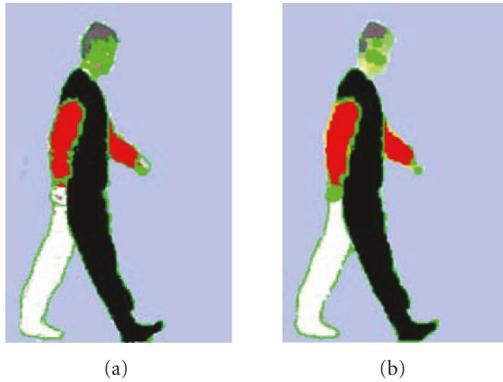


FIGURE 4: (a) Color segmentation before dilation and erosion, (b) resultant image after dilation and erosion.

4.1. Color segmentation

Although the previous work to segment images in the grayscale domain produced promising results, fine adjustments were required to improve the image quality every time a new image was processed. In addition, the image acquisition subsystem was able to capture high-quality color images, processing those images in grayscale domain has wasted the rich information carried in the color images. Hence, color domain processing is investigated in this work. Furthermore, because color segmentation is more sensitive to the change of color, this approach gives us more flexibility in the costume design (less distinct colors can be used). Hence, the capability of the image processing subsystem will eventually be improved. In this paper, the RGB color format is chosen, as illustrated in Figure 3.

Segmentation of the grayscale images was done by threshold technique to trace the boundary of each segment. Subsequently, a region growth algorithm is applied to each enclosed region to complete the segmentation process. To yield an optimal outcome, the contrast of each region needs to be significant. As a result, image qualities such as brightness and contrast need to be adjusted manually for every new image. Imagine the same procedure is applied to three different layers for RGB image format. It would be even harder to obtain a good color-segmentation. Further more, as observed from Figure 3, it would be ineffective by human eyes to distinguish red and black in both green and blue layers, let alone by computer programs. Hence, a neural network approach for color segmentation is proposed.

The RGB values of various pixels from different regions were used to train a back propagation neural network to guide the system to recognize different colors presented in the image. The network consists of three inputs corresponding to the RGB values in every pixel. Four outputs were used to represent the four colors appeared in the image, corresponding to *red*—the upper limb color, *black*—the torso color, *white*—the back leg color, and *blue*—the color of the backdrop. The network was trained by taking small patches of pixels from each region, hence learning to recognize the color in different regions. Then, the trained

network is applied to separate different regions in the images. Once the network has identified the color represented in each pixel, it then produces a segmentation map according to the outcome of each pixel. Figure 4(a) shows an example of the segmentation result in the RGB color space with red = (255,0,0), black = (0,0,0), white = (255,255,255), and blue = (175,190,240). This was to ensure that the color of each region was uniformly presented in the image. Pixels failed to be classified successfully will remain its original RGB values. Morphological processing techniques [21] are then applied to remove the misclassified small regions. Dilation has the effect of expanding an image whereas erosion has the opposite effect. Different combinations of dilation and erosion could be used to remove the small regions effectively. It is experimentally determined that the best sequence of the dilation and erosion operations was *Dilation* → *Dilation* → *Erosion* → *Erosion*, as illustrated in Figure 4(b). Note that although the segmentation of the head is still not satisfactory, it had little effect in our current gait analysis.

A boundary line has also been traced between different regions. The process was done by scanning the pixels of the segmentation map horizontally. At each point, the value of the previous pixel and next pixel was compared, if the values were not the same, that is, there was a color change taking place. The original color of the current pixel was replaced by a green dot, and then move to the next pixel. Figure 5 illustrates the flowchart of the process.

The boundary outlines the shape of each region, and the medial point can be determined by taking the average of the two boundary points at every horizontal cross section. The resultant color segmentation is shown in Figure 4(b), which consists of one torso segment (black), two arm segments (red), and one back leg segment (white).

The resultant image after the dilation/erosion processes removes the unclassified points on the face and the hand. It also fills the impurity at the edge of the white leg (see Figure 4(a)) to form a complete region (see Figure 4(b)).

The body segment is incomplete due to the fact that the front arm obscures part of the body. Therefore, further processing of this segment is required, such that the skeletonization procedure can be properly carried out on the body segment. The body restoration process is adopted from the algorithm developed in [15]. A reference image, which shows that the position of the front arm is located within the torso region, is chosen. This method assumes that the position of the upper body, in terms of tilt or stoop, does not vary much through the gait sequence. Figures 6(a) and 6(b) display the original and reference images. Figures 6(c) and 6(d) represent the body segment before and after restoration.

4.2. Skeletonization

Following the segmentation of the image, we are ready to proceed with the extraction of the relevant representation of the subjects. Clinical experience shows that these features are joint angles and swing distances. Extracting these features is made easier by skeletonize the segmented image. The skeletons in this paper are assumed to be the medial axis of each particular segment found in segmentation. The first

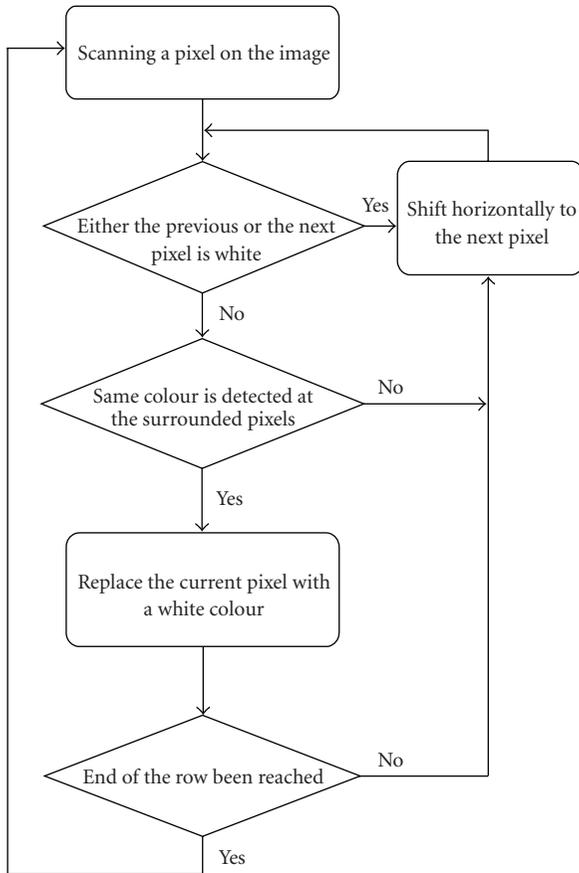


FIGURE 5: Flow chart for boundary detection.

task is to thin each of the segments obtained earlier. The process is done by adopting the algorithm developed in [22]. The algorithm performs the medial axis transform on each body segment separately. Each thinned body segment is then further processed to remove all the erroneous branches in the partial skeleton, which were caused by portions of segment protruding from the proper image segment. This is important especially when it comes to the body segment. In some cases, a small protrusion is left as a remnant after the body segment restoration. This is enough for the thinning algorithm to see it as a separate part to the segment that needs to be thinned. Thus, a branch will be created out from the central skeleton of the segment, as illustrated in Figure 7(a). The skeletons of each segment are then combined to give the whole body skeleton. This skeleton is used to measure the arm swing distances. Figure 7(a) represents a skeleton that was combined without branch removal, as compared to the clean whole body skeleton shown in Figure 7(b).

5. FEATURE EXTRACTION

Neurological signs in PD patients are typically characterized by symptoms such as tremor, bradykinesia and rigidity, and gait/posture stability. These symptoms can show significant differences between normal people and PD patients in terms of the gait features, such as joint angles, swing distances,

and swing trajectories of the limbs which can be analyzed from both static and dynamic perspectives. This research focuses on analyzing Parkinsonism from the static processing perspective due to its simplicity and relative effectiveness. Specifically, two groups of features will be extracted: (1) swing distances between the ends of the limbs and those between the ends of the limbs and the median axis of the torso, (2) joint angles between sections of the limbs and those between the limbs and the torso. The extracted features will be analyzed and used as the inputs to the decision making subsystem for the classification.

In this work, four features were considered in the distance group and six in the angle group. These features are illustrated in Figure 8.

The four features in the distance group are

- (i) front hand to the median axis of torso. (F8),
- (ii) back hand to the median axis of torso. (F9),
- (iii) front hand to back hand (F10),
- (iv) heel of the front foot to the toe of the back foot (F7).

The value of the arm swing distance may be negative to indicate the swing direction. The purpose is to distinguish the abnormality of the arm swing of the patients from the normal people. To keep the measurement uniform amongst different samples, all the distance features are normalized by the height of the human subject.

In the joint angle group, six features are extracted:

- (i) two knee-joint angles (F1 and F4),
- (ii) two ankle-joint angles (F2 and F3),
- (iii) two joint angles at the elbows (F5 and F6).

5.1. Distance feature calculation

It is noted that the swing distance of the limb is the most suitable feature to assess the ability of the limbs to stretch (stretchability). It is also apparent that there are certain correlations between the distances and the joint angles, which makes it feasible to substitute the distance for the joint of the same limb. For example, when the arms fully swing, both the swing distance and the shoulder joint angle will reach their maximum values simultaneously. The best option to represent the flexibility of the limb would be the joint angle features. However, it is in general difficult to obtain precise angle measurement. We therefore also consider using distance features.

The advantages of using the distance features are multitude. First of all, they are relatively easier to obtain. It only requires two end points to extract this feature. Secondly, the results of the distance calculation are robust to noise, due to the fact that the distance features calculation only depends on the two end points. In general, the obtained distance features are more accurate than the angles features calculated by the current method based on skeleton.

The distances measurement is based on the skeleton figure obtained, since it can accurately represent the general structure of the human subject. We identify the end points of

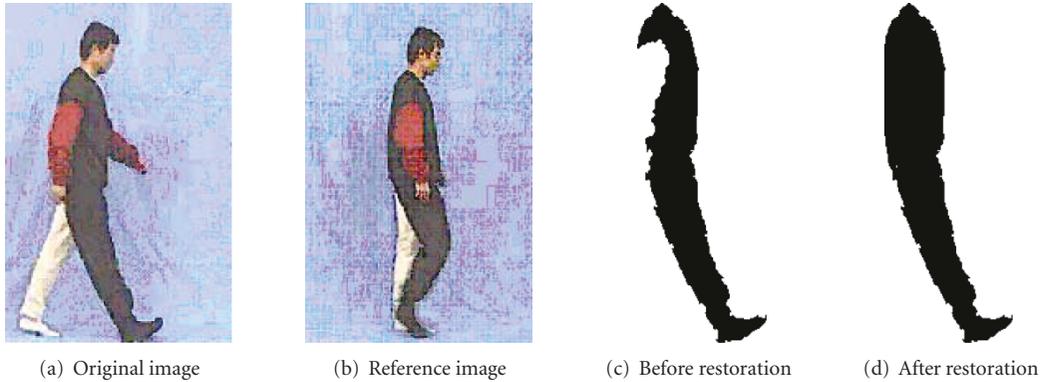


FIGURE 6: Image restoration example.

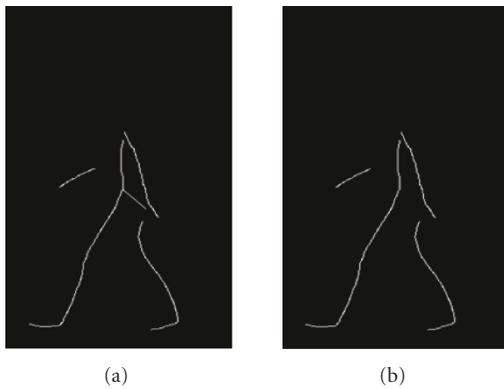


FIGURE 7: Skeletonization examples of (a) whole body skeleton, and (b) skeleton after debranch.

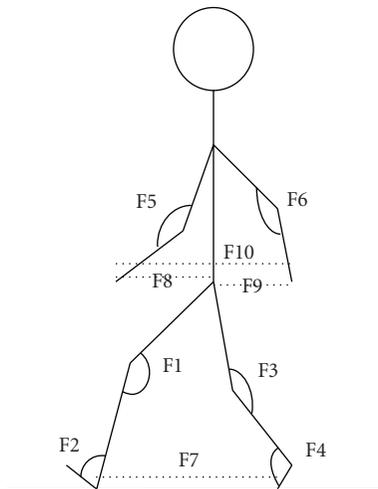


FIGURE 8: Locations of all features.

the limbs from the skeleton, and then calculate the distances. First, the center of the torso in the horizontal direction needs to be found. This is not as easy as it sounds, because of the variations in the physical build of the human subjects can be substantial. Particularly in the aged group, the median point

of the body around the chest in the horizontal direction is not very accurate. Through the analysis of the data samples, it is apparent that the two legs' merging point around the hip area more accurately represents the real median point of the body. Also it is noted that the bearings of the gait are crucial while extracting the distance between the front foot heel and the back foot toe (F7). In this project, the black leg is always in the front, and the white leg is away from the camera. Since the walking direction can be either to the right or to the left, the walking direction for each individual image has to be identified before any further processing. For the distance between the two feet, a chain coding algorithm is used [22]. The measurement of the distance for the arm is based on the skeleton figure to the arm region only, because it identifies the end points from the skeleton of the arm.

5.2. Joint angle calculation

Previous work attempted to use chain coding to represent the skeleton figure, and from the directional information of the chain code, to calculate the various joint angles. This method was found to be very sensitive to straight line segments. Therefore, a more robust method based on the Hough transform is investigated.

5.2.1. Hough transform

The Hough transform [1] transforms an image into the parameter space, where the image is represented by the parameters of straight lines or curves inherent in the original image plane. One of the most popular applications of Hough transform is edge linking, where the objective is to link together separate segments of lines in an image [23]. This involves finding the subsets of points, which lie on the same straight lines.

Consider a point (x_i, y_i) on a straight line in an image. The Hough transform can take two forms to represent the straight line as the input, one in the slope-intercept form, $y_i = ax_i + b$, and the other in the normal form, $x \cos \theta + y \sin \theta = \rho$. The drawback of using the slope-intercept form is that the gradient becomes infinity for vertical lines. Therefore, the normal representation of a line is the preferred

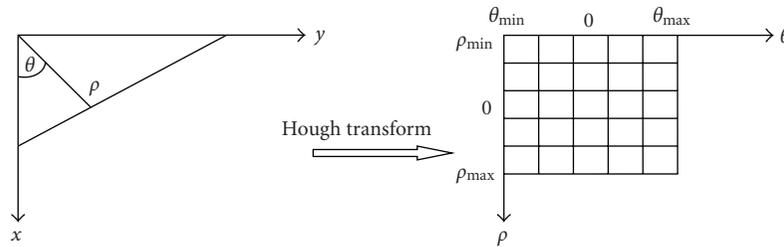


FIGURE 9: Hough transform using normal representation of a line.

form. This is illustrated graphically in Figure 9. Since a computer processes information digitally, the original image is quantised spatially. Similarly, the Hough transformed image must also be quantized. This is just a matter of using a 2D matrix with a desired number of elements for θ and ρ . In Figure 9, the transformed image has a θ_{\min} of -90 degrees, θ_{\max} of $+90$ degrees, ρ_{\min} of $-\sqrt{2}D$ pixels, and ρ_{\max} of $+\sqrt{2}D$ pixels, where D is the distance between the corners of the original image in pixels. The actual number of elements in the parameter space depends on the desired accuracy and resolution. The larger the number of θ cells, the finer the resolution of the angle of a single line is. The larger the number of ρ cells, the straighter the line represented by one cell of the transformed image. Each cell contains the number of pixels in the image that lie on the line with the corresponding parameter values.

5.2.2. Joint angles extraction

The knee joint angle is the angle between the upper and lower legs. This angle is always more than 90 degrees for walking sequences, since a person have to be running to achieve knee angle of less than 90 degrees. Once the different limbs have been extracted, the Hough transform is applied to obtain their parametric representations. The cell with the highest pixel counts in the parameter matrix then represents a straight line with the most pixels through it and thus represents the most likely line through that part of the skeleton. The absolute angle of this line can be found immediately by reading the θ coordinate of the cell. This process is repeated until the straight lines for all the limbs are found.

5.3. Quantitative evaluation of the precision of the features

Both the distance and the angle features are estimated based on the stick figure calculated by skeletonization. When the image is noise free, the stick figure is the medial axis of the human body (with subpixel precision), and the accuracy of the distances is also within subpixel precision. The precision of estimating the joint angles depends on the resolution of the θ cells. The larger the number of θ cells, the finer the resolution of the joint angle one can obtain, but the more computationally intensive the algorithm becomes. In this work, the resolution of the angle is approximately 0.5

degree which is a proper compromise between accuracy and efficiency.

6. FEATURE SELECTION

Previous work attempted to use histogram analysis to select the most significant features to train a multilayer back propagation neural network [15]. Although it achieved a classification accuracy of approximately 85%, it requires a subjective judgment of the result to obtain a reliable feature selection. Therefore, we investigated a hybrid feature selection method combining the strengths of a sequential backward selection (SBS) procedure for feature selection and a general regression neural network (GRNN) for feature evaluation.

6.1. General regression neural networks (GRNN)

GRNN can be considered as a special example of the radial basis function (RBF) network, where the units in the hidden layer adopt the Gaussian kernel as the non-linear activation function while the second layer consists of linear summation units. Unlike the conventional RBF network where the centers and the widths of the Gaussian kernels are determined by iterative clustering procedures, the corresponding parameters in GRNN are represented as a deterministic function of the training data. In other words, no iterative training procedure is required to reconstruct a mapping using GRNN, hence allowing rapid evaluation of the relevancy of different feature subsets [18, 24].

6.1.1. Mathematical background of GRNN

GRNN is a memory-based feed-forward neural network. The regression of a dependent variable, Y , on an independent variable, X , is the computation of the most probable value of Y for each value of X based on a finite number of possibly noisy measurements of X and the associated values of Y . The variables X and Y are usually vectors. In system identification, the dependent variable, Y , is the system output, and the independent variable, X , is the system input. In order to implement system identification, it is usually necessary to assume some functional form with unknown parameters a_j . The values of the parameters are chosen to make the best fit to the observed data. In the case of pattern recognition, the independent variable X denotes the feature vector of the pattern to be classified, and Y is

the classification result. Assume that $f(x, y)$ represents the known joint continuous probability density function of a vector random variable, x , and a scalar random variable, y . Let X denote a particular measured value of the random variable x . The conditional mean of y given X is

$$E[y|X] = \frac{\int_{-\infty}^{\infty} y f(X, y) dy}{\int_{-\infty}^{\infty} f(X, y) dy}, \quad (1)$$

where the density $f(x, y)$ is usually unknown, and in GRNN, this probability density function is usually estimated from samples of observations of x and y using nonparametric estimators. The estimator used in this work is the class of consistent estimators proposed by Parzen [18]. This probability estimator $\hat{f}(X, Y)$ is based upon sample values X_i and Y_i of the random variables x and y :

$$\begin{aligned} \hat{f}(X, Y) &= \frac{1}{(2\pi)^{(p+1)/2} \sigma^{p+1}} \\ &\times \frac{1}{n} \sum_{i=1}^n \exp \left[-\frac{(X - X_i)^T (X - X_i)}{2\sigma^2} \right] \\ &\times \exp \left[-\frac{(Y - Y_i)^2}{2\sigma^2} \right], \end{aligned} \quad (2)$$

where n denotes the number of samples, and p denotes the dimension of the vector variable x . A physical interpretation of the probability estimator $\hat{f}(X, Y)$ is that it assigns sample probability of width σ for each sample X_i and Y_i , and the probability estimate is the sum of those sample probabilities. Substituting the joint probability estimate $\hat{f}(X, Y)$ into the conditional mean yields

$$\hat{Y}(X) = \frac{\sum_{i=1}^n Y_i \exp(-D_i^2/2\sigma^2)}{\sum_{i=1}^n \exp(-D_i^2/2\sigma^2)}, \quad (3)$$

where D_i^2 is defined as

$$D_i^2 = (X - X_i)^T (X - X_i). \quad (4)$$

The only known parameter in the above equation is the width of the estimating kernel which can be estimated by using a cross validation method called the leave-one-out method. For a particular value of σ with a training data set of n samples, the leave-one-out method moves one sample at a time and constructs the GRNN using the remaining $(n - 1)$ samples. Then, the GRNN is used to classify the sample excluded. This is repeated n times, and each classification result is stored. Then, the mean square classification error σ is calculated.

6.1.2. General regression neural network architecture

The above-mentioned regression algorithm can be implemented in a neural network architecture which is shown in Figure 10. It consists of four layers: the input layer, the hidden layer, the summation layer, and the output layer. The function of the input layer is to pass the input vector

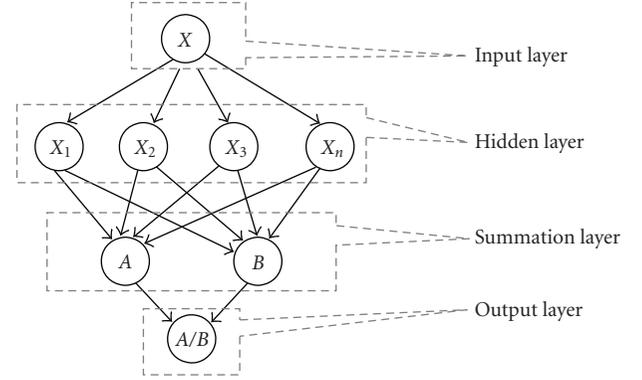


FIGURE 10: GRNN architecture.

variables X to all the units in the hidden layer. The hidden layer consists of all the training samples X_1, \dots, X_n . When an unknown pattern X is presented, the squared distance D_i^2 between the unknown pattern and the training sample is calculated and passed through the kernel function. The summation layer has two units A and B , unit A computes the summation of $\exp[-D_i^2/(2\sigma^2)]$ multiplied by the Y_i associated with X_i . The B unit computes the summation of $\exp[-D_i^2/(2\sigma^2)]$. The output unit divides A by B to provide the prediction result.

6.2. The feature selection procedure

The walking sequences of 90 people were video-taped and processed, including 50 from normal people (the control group) and 40 from the patients group. These samples were evenly divided into four subgroups: NA, NB, PA, and PB, where N represents normal people, P represents patients, A represents the training data, and B represents the testing data. Various techniques were carried out to select the best features for classification.

- (i) Features selected by color histograms.
- (ii) Features selected by SBS/GRNN.
- (iii) Features selected by the combinations of color histogram and SBS/GRNN.
- (iv) All the features calculated in feature extraction.

6.2.1. Features selected from histogram analysis

In the histogram analysis technique, ten color histogram features are obtained. The features with the best classification power between the control group and the patient group had histograms showing little overlap between the normal and patient data. On the other hand, the color histograms of features displaying substantial overlap perform poorly in classification (see Figure 11). We observed that the most distinguished features obtained then are the front arm swing distance, the front elbow angle, back foot's ankle angle, back arm swing distance, and from leg angle at the knee [25].

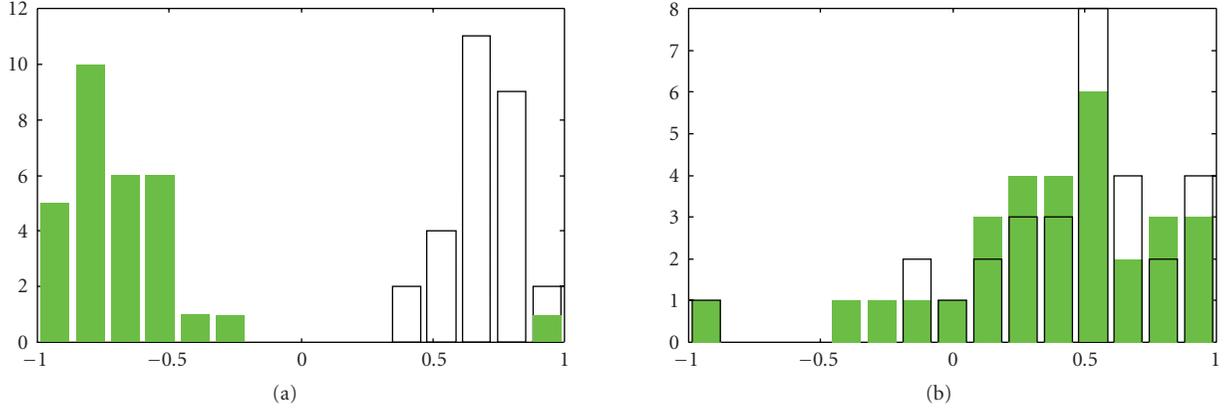


FIGURE 11: Examples of (a) good and (b) bad feature. The green area represents the patient group, and the blank area represents the control group.

TABLE 1: The order of the discriminatory power of the 10 features using the SBS method.

Order	Feature name:
1	Front foot ankle angle.
2	Back hand to median axis of torso.
3	Front hand to back hand.
4	Heel of the front foot to the toe of the back foot.
5	Back knee joint angle.
6	Front elbow angle.
7	Back foot ankle angle.
8	Front knee joint angle.
9	Back elbow angle.
10	Front hand to median axis of torso.

6.2.2. Feature selection by SBS/GRNN

The sequential backward selection (SBS) method is a simple top-down feature selection procedure. Starting from the complete set of features, Y features are discarded one by one at each stage. The feature discarded is the one with the least discriminatory power from the current feature set. Assume k features have been discarded to form a feature set, X_k , which has $N - k$ features at this moment. Then, feature F_j is discarded from the remaining $N - k$ features if

$$\sigma(X_k \cup F_j) \geq \sigma(X_k \cup F_i), \quad \text{for } i = 1, 2, \dots, N - k, i \neq j, \quad (5)$$

where σ is the mean squares error function defined before, evaluating classification performance. Then, $X_{k+1} = X_k - F_j$. The Algorithm is initialized by setting $X_0 = Y$.

To identify the most discriminatory feature set, the algorithm will stop at a point where the classification error begins to increase if removing more features from the current feature set. To find the relative importance of all the N features within the set Y , the algorithm will continue until only one feature remains.

The SBS method is first used to find out the discriminatory power of those 10 features, described in the previous

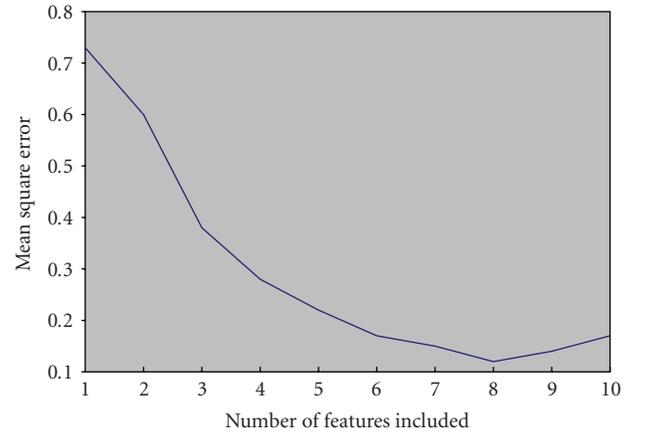


FIGURE 12: The result of sequential backward selection method.

chapter. The order of the discriminatory power of those 10 features is shown in Table 1. The mean square versus the number of feature included in classification is plotted in Figure 12. We observe that the minimum mean square error occurs when the top eight discriminatory features are included.

As noted in the resulting graph for the SBS algorithms (see Figure 12), the error curves exhibit a distinct minimum point at the feature subset X_8 and beyond which the error starts to increase again. Since the GRNN modeling process does not incorporate any explicit trainable parameters, it would be difficult for GRNN to model the characteristics of the features beyond a certain maximum number of features. For limited number of training samples and their increasing sparseness in high-dimensional spaces, the error starts to rise beyond the local minimum point.

6.2.3. Combination of histogram and SBS/GRNN

Although the discriminatory power given in Table 2 is supposed to indicate the importance of the individual features in characterizing normal and patient groups, this

TABLE 2: Classification result.

	Experiment 1	Experiment 2	Experiment 3	Experiment 4	Average
Histogram	83.8%	84.1%	82.4%	84.8%	83.7%
SBS/GRNN	84.3%	85.1%	84.2%	85.6%	84.8%
Histogram and SBS/GRNN	88.3%	89.1%	88.0%	88.2%	88.4%
All features	75.1%	76.4%	76.0%	77.1%	76.1%

may not be the case for those features beyond the minimum point. Hence, a technique which combines SBS/GRNN and histogram was proposed. It obtains five good features using GRNN technique, and then uses the histogram to select the two features with the most discriminatory power from the remaining four features.

- (i) The resulting five features from the SBS algorithm are front hand to back hand distance, front elbow angle, back elbow angle, heel of the front foot to the toe of the back foot distance, and back ankle angle.
- (ii) The best two of the remaining features obtained by histogram analysis are back knee joint angle and front hand to torso distance.

7. CLASSIFICATION RESULTS

We implemented the proposed processing method into our video analysis system and tested the system on a database of both PD patients (the patient group) and healthy people (the control group). The dataset was video-taped at Sydney Westmead Hospital, and at the University of Sydney. In total, 90 images (40 from the patients group and 50 from the normal group) were extracted from the videos and processed. The samples from the two groups were evenly divided into four subgroups: NA, NB, PA, and PB (N stands for normal and P for patient). The data were divided in such fashion, hence while using group A (NA and PA) as training data, group B was used as testing data, and vice versa.

The *decision making part* consists of a three-layered back propagation neural network. It was trained by the features selected from the previous stage, and the performance of the system was tested using new images. Separate neural networks were constructed for training features selected by the above techniques.

- (1) Histogram analysis. For the features obtained by histogram technique, a network consisting of 6 input neurons corresponding to the six most significant features selected by histogram analysis, namely, back knee angle (F4), both elbow angles (F5, F6), front hand swing distance (F8), back hand swing distance (F9) and the distance between the front hand and back hand (F10), 4 hidden neurons, and 2 output neurons was constructed. The network was trained using the six features selected by the histogram analysis.
- (2) Sequential backward selection. The second network consists of 8 input neurons, 5 hidden neurons, and 2 output neurons. The eight most effective features

determined by the SBS algorithm were used to train this network.

- (3) Combined SBS and histogram analysis. This network consists of 7 input neurons; each corresponds to the features obtained by this feature. The network consists of 6 hidden neurons and 2 output neurons. The network uses five of the most significant features selected using SBS/GRNN and two other features, backhand elbow angle (F6) and backhand swing distance (F8), obtained from histogram analysis as inputs to train the network.
- (4) All the features included. The last network was constructed to be trained with all 10 features and compared to the classification results obtained by the feature extraction strategies described above. Six hidden neurons and 2 output neurons were implemented in this network.

Four experiments were conducted: (1) using NA, PA as the training set and NB, PB as the testing set, (2) using NB, PB as the training set and NA, PA as the testing set, (3) using NA, PB as the training set and NB, PA as the testing set, (4) using NB, PA for training and NA, PB for testing. The data were subdivided into those four groups to avoid over training in the case of a large sample space. Those four experiments were performed for cross validation.

From Table 2, we have the following observations.

- (i) The histogram analysis method selects the features according to the data distribution for each feature. By discarding the “bad features”, the system uses only “good” and “medium” for the classification. The system yields an 83.7% correct detection rate which is acceptable.
- (ii) Selecting features by SBS/GRNN, the classification results were slightly improved compared with those obtained from histogram analysis. This indicates that SBS/GRNN, the more systematic feature selection method, is able to better explore the information carried in the database in terms of the features being studied.
- (iii) By combining the best features selected by SBS/GRNN and histogram analysis, the average classification rate has been further improved to 88.4%. This clearly demonstrates the complementary discriminatory powers of the two feature selection methods.
- (iv) When all the features were used for classification, the results were substantially worse than those obtained using the selected sets of features.

In conclusion, there is not enough evidence to justify whether distance features are better than joint angle features in determining Parkinsonism patients. However, from the results, it is clear that certain features, such as front arm swing distance and front elbow angle, are more significant in classifying Parkinson symptom than others. When all features, including those features with less significance, were used for classification, the results clearly dropped.

8. CONCLUSIONS AND FUTURE WORK

This paper applies image processing techniques to assist clinical analysis of patients with gait altering neurological disorders. The main objective of the work is to remove constraints of laboratory environment and to provide a more realistic setup for assist in diagnosing neurological disorders. In the paper, a neural network based color image segmentation method was introduced to effectively utilize the rich information in the color domain to improve quality of object segmentation, and a feature selection method-based sequential backward selection and general regression neural network was combined with a histogram feature selection mechanism to select an effective feature set for robust decision making.

The system developed in this work is able to automatically extract features, swing distances, and joint angles, under the given environment setting. Hence, it provides a solid foundation for developing intelligent computer-assisted systems to assist neurologists in diagnosing posture and movement disorders.

The image acquisition system in this work still needs some constrained settings. The reason to have the setting is due to the difficulty of image segmentation. As we all know, state of the art in image processing is still not able to support high quality segmentation in a complete setup free environment. However, new techniques in interactive computer vision such as graph cut [26] have demonstrated their effectiveness in segmenting objects in relatively complex environment in an interactive fashion, thus providing a potential solution to image segmentation when real-time processing is not required as in the application on hand. We will investigate application of graph cut to video analysis of human gait and posture for assisting clinicians in diagnosing Parkinson disease.

REFERENCES

- [1] D. H. Ballard and C. M. Brown, *Computer Vision*, Prentice-Hall, Englewood-Cliffs, NJ, USA, 1982.
- [2] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.
- [3] M. Selzer, S. Clarke, L. Cohen, P. Duncan, and F. Gage, *Textbook of Neural Repair and Rehabilitation*, Cambridge University Press, Cambridge, UK, 2006.
- [4] C. K. William and R. L. Watts, *Movement Disorders: Neurologic Principle & Practice*, McGraw-Hill Professional, New York, NY, USA, 2004.
- [5] A. C. England and R. S. Schwab, "Parkinson's syndrome," *The New England Journal of Medicine*, vol. 265, pp. 785–887, 1961.
- [6] J. M. S. Pearce, *Parkinson's Disease and Its Management*, Oxford University Press, New York, NY, USA, 1992.
- [7] L. A. Gundersen, D. R. Valle, A. E. Barr, J. V. Danoff, S. J. Stanhope, and L. Snyder-Mackler, "Bilateral analysis of the knee and ankle during gait: an examination of the relationship between lateral dominance and symmetry," *Physical Therapy*, vol. 69, no. 8, pp. 640–650, 1989.
- [8] J. Han, H. S. Jeon, B. S. Jeon, and K. S. Park, "Gait detection from three dimensional acceleration signals of ankles for patients with Parkinson's disease," in *Proceedings of the International Special Topic Conference on Information Technology in Biomedicine*, Ioannina, Epirus, Greece, October 2006.
- [9] L. Lee and W. E. L. Grimson, "Gait analysis for recognition and classification," in *Proceedings of the 5th IEEE International Conference on Automatic Face and Gesture Recognition (FGR '02)*, pp. 148–155, Washington, DC, USA, May 2002.
- [10] H. Mitoma, R. Hayashi, N. Yanagisawa, and H. Tsukagoshi, "Characteristics of parkinsonian and ataxic gaits: a study using surface electromyograms, angular displacements and floor reaction forces," *Journal of the Neurological Sciences*, vol. 174, no. 1, pp. 22–39, 2000.
- [11] M. W. Whittle, *Gait Analysis: An Introduction*, Elsevier Health Sciences, Oxford, UK, 2002.
- [12] R. Chang, L. Guan, and J. A. Burne, "An automated form of video image analysis applied to classification of movement disorders," *Disability and Rehabilitation*, vol. 22, no. 1-2, pp. 97–108, 2000.
- [13] R. D. Green and L. Guan, "Quantifying and recognizing human movement patterns from monocular video images—part II: applications to biometrics," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 2, pp. 191–198, 2004.
- [14] R. D. Green, L. Guan, and J. A. Burne, "Real-time gait analysis for diagnosing movement disorders," *Journal of Electronic Imaging*, vol. 5, pp. 253–269, 1999.
- [15] T. Tan, L. Guan, and J. A. Burne, "A real-time image analysis system for computer-assisted diagnosis of neurological disorders," *Real-Time Imaging*, vol. 5, no. 4, pp. 253–269, 1999.
- [16] L. Wang, T. Tan, W. Hu, and H. Ning, "Automatic gait recognition based on statistical shape analysis," *IEEE Transactions on Image Processing*, vol. 12, no. 9, pp. 1120–1131, 2003.
- [17] C.-Y. Yam, M. S. Nixon, and J. N. Carter, "Extended model-based automatic gait recognition of walking and running," in *Proceedings of the 3rd International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA '01)*, pp. 278–283, Halmstad, Sweden, June 2001.
- [18] M. S. Nixon and J. N. Carter, "Advances in automatic gait recognition," in *Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition (FGR '04)*, pp. 139–144, Seoul, Korea, May 2004.
- [19] M. S. Nixon and J. N. Carter, "Automatic recognition by gait," *Proceedings of the IEEE*, vol. 94, no. 11, pp. 2013–2024, 2006.
- [20] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer, "The humanID gait challenge problem: data sets, performance, and analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 2, pp. 162–177, 2005.
- [21] A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall, Englewood-Cliffs, NJ, USA, 1989.

- [22] R. Chang, L. Guan, and J. A. Burne, "A computer assisted image analysis system for diagnosing movement disorders," in *Proceedings of the 10th Australian Joint Conference on Artificial Intelligence (AI '97)*, pp. 290–301, Perth, Australia, November–December 1997.
- [23] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison-Wesley, Reading, Mass, USA, 1993.
- [24] D. F. Specht, "A general regression neural network," *IEEE Transactions on Neural Networks*, vol. 2, no. 6, pp. 568–576, 1991.
- [25] H. Lee, "Video analysis of human gait and posture to determine neurological disorders," Master of Engineering Thesis, University of Sydney, Sydney, Australia, August 2000.
- [26] Y. Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images," in *Proceedings of the 8th IEEE International Conference on Computer Vision (ICCV '01)*, vol. 1, pp. 105–112, Vancouver, BC, Canada, July 2001.