Computers in Biology and Medicine 42 (2012) 925–934

Contents lists available at SciVerse ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Analysis of adductors angle measurement in Hammersmith infant neurological examinations using mean shift segmentation and feature point based object tracking

D.P. Dogra [a,*], A.K. Majumdar [a], S. Sural [b], J. Mukherjee [a], S. Mukherjee [c], A. Singh [c]

[a] Department of Computer Science & Engineering, Indian Institute of Technology, Kharagpur 721302, India
[b] School of Information Technology, Indian Institute of Technology, Kharagpur 721302, India
[c] Institute of Post Graduate Medical Education and Research, SSKM Hospital, Kolkata 700020, India

ARTICLE INFO

Article history:
Received 3 July 2011
Accepted 8 July 2012

Keywords:
Infant neurological examinations
Adductors angle measurement
Image segmentation
Thinning
Object tracking
Video analysis
Straight line fitting
Exponential regression

ABSTRACT

This paper presents image and video analysis based schemes to automate the process of adductors angle measurement which is carried out on infants as a part of Hammersmith Infant Neurological Examination (HINE). Image segmentation, thinning and feature point based object tracking are used for automating the analysis. Segmentation outputs are processed with a novel region merging algorithm. It is found that the refined segmentation outputs can successfully be used to extract features in the context of the application under consideration. Next, a heuristic based filtering algorithm is applied on the thinned structures for locating necessary points to measure adductors angle. A semi-automatic scheme based on the object tracking of a video has been proposed to minimize errors of the image based analysis. It is observed that the video-based analysis outperforms the image-based method. A fully automatic method has also been proposed and compared with the semi-automatic algorithm. The proposed methods have been tested with several videos recorded from hospitals and the results have been found to be satisfactory in the present context.

1. Introduction

Video and image analysis can be used to diagnose medical problems even with new born babies. According to medical experts, a significant proportion of highly preterm infants (gestational age at birth less than 32 weeks) usually show neurological and developmental disabilities [1]. Due to Intra-ventricular Haemorrhage or Peri-ventricular Leukomalacia, an increasing prevalence of Cerebral Palsy can occur in premature, low birth weight newborns and children born with asphyxia [2]. It has been reported that early diagnosis can increase the survival rate of patients with very low birth weight and neurological disorders [3]. A popular method to diagnose such disorders in infants is known as Hammersmith Infant Neurological Examination (HINE) [4]. It is a quantitative method for assessing neurological development of infants between 2 and 24 months of gestational age. The examinations include assessment of cranial nerve functions, posture, movements, tone, reflexes, and behavior. An overall score quantifying the neurological development index of the infant is estimated from the observations made during the above mentioned examinations.

It has been learnt from domain experts that the conventional way of HINE is not user friendly. Even a highly experienced person may find some of the examinations difficult to conduct. The motivation behind this work came from the fact that image and video processing schemes can be used successfully to aid the physicians for conducting HINE tests. Though computerized techniques for infant tele-monitoring [5], baby gesture recognition [6], and infant behavior analysis [7] are available, no work related to automatic analysis of neurological examinations following HINE procedure [4] has been reported. Presently, doctors completely rely on visual observation to assess the neuro-physiological growth of a child.

We have addressed some of the problems associated with the conventional HINE methodology and the results have been published recently. For example, in [8,9], we have discussed that videos and images of HINE tests can be recorded with minimal interference to the conventional examination setup. In [10], it has been shown that, using feature point based video object tracking and with the help of a 2D geometrical model, analysis of HINE pulled-to-sit examination can be automated. However, there are a few angle measurement based examinations which have not been covered yet. HINE adductors angle measurement is one such

* Corresponding author. Tel.: +91 97 33 95 8129.
E-mail addresses: debiprosadcom@yahoo.com (D.P. Dogra), akmj@cse.iitkgp.ernet.in (A.K. Majumdar), shamik@sit.iitkgp.ernet.in (S. Sural), jay@cse.iitkgp.ernet.in (J. Mukherjee).

0010-4825/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compbiomed.2012.07.005

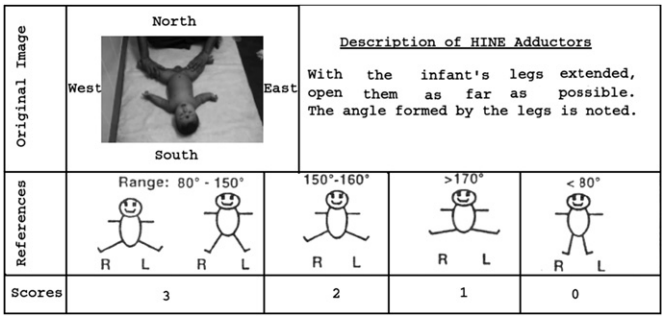

Fig. 1. Sample image of conventional adductors angle measurement with reference templates and scores. The original image was captured in the neuro-development clinic of Institute of Post Graduate Medical Education & Research (IPGMER) and Seth Sukhlal Karnani Memorial (SSKM) Hospital, Kolkata, India.

examination. The process is illustrated in Fig. 1. While conducting these tests, the physician places a Goniometer [11] across the extended legs of the infant after putting the baby at supine posture. At the beginning, the examiner stretches the legs of the baby and when the first resistance from the baby is experienced, the measurement of angle is recorded. Depending on the measured angle, a score as mentioned in Fig. 1 is assigned. Though it is currently used in practice, the scheme is not user friendly and often the physicians find it difficult to keep the baby in stable position. Some of the motivating factors behind developing assistive tools to help the physicians are discussed below.

- Offline analysis is one of the advantages of video and image based schemes. In the conventional approach, the examiner assesses the neurological development instantly. Repeating the same examination multiple times may increase the discomfort level of the infant undergoing the assessment process. However, measurements can be repeated if visual recordings of the examinations are available.
- It has been observed that angle measurement using Goniometer [11] can be inconvenient and the method is error prone due to movement of the baby undergoing examinations. In such situations, image or video based analysis can be advantageous.
- The conventional HINE process is solely based on visual observation. That is why assessment of body movements often turns out to be difficult.

This paper proposes a method to automate the analysis of HINE adductors angle estimation using image and video processing based algorithms. It has been discussed in this paper that how image segmentation and object tracking schemes can be successfully used to help the physicians for analyzing HINE adductors angle measurement videos. First, an image based analysis of the examination has been carried out using a context based feature extraction method. Preliminary results of the image based analysis have been reported in [12]. However, it has been observed that single image based scheme may fail due to the errors of the intermediate steps of the feature extraction process. Thus, an alternative approach using video object tracking has been proposed. It is found that the proposed video based analysis can overcome the limitations of the image based analysis.

Rest of the paper is organized as follows. In Section 2, feature extraction steps of the image based analysis have been reported. In the next section, a methodology to analyze adductors angle measurement using image based features is presented. Next, in Section 4, the proposed video based analysis using object tracking is discussed. Comparative results of the schemes are presented in Section 5. Conclusion and future directions of the present work are reported in Section 6.

2. Feature extraction

Results using the feature extraction process reported in [12] have been improved in this paper using a novel region merging scheme. First, an image segmentation based background/foreground separation technique is applied to identify the segments where the baby is located. Next, a post-processing algorithm is proposed for refining the segmentation outputs. To identify features required for adductors angle measurement, binarization and thinning of the foreground object (corresponding to the identified segment where the baby is located) have been carried out. The heuristic proposed in [12] has been used to filter out noisy edges that may be generated during the thinning process. Finally, the filtered skeleton images are used to extract features for angle measurement.

2.1. Segmentation and object identification

We have experimented with several image segmentation schemes that are commonly used for partitioning natural images [13–16]. It has been found that mean-shift filtering based segmentation proposed by Comaniciu et al. [13] outperforms other schemes when used for segmenting images of HINE adductors examination [12]. The effectiveness of mean-shift based algorithm has also been established by Unnikrishnan et al. [17]. However, the segmentation parameters need to be tuned for the class of images being processed. After carrying out several experiments with the images of this domain, it has been found that with spatial bandwidth $(h_s) = 7$, range bandwidth $(h_r) = 12.5$, and minimum region size 300 pixels, mean-shift outperforms other algorithms.

One of the segmentation outputs is shown in Fig. 2(b). It is evident from the outputs that the foreground object (the baby undergoing examination) can be separated from the background using skin color information. However, a single region may not cover the entire foreground object. Small and insignificant regions may exist which usually makes the feature extraction process erroneous. To overcome this, a region merging algorithm has been proposed that successfully merges insignificant regions with neighboring segments having similar color and texture.

2.2. Post-processing using region merging

After initial classification of the pixels using color based segmentation, a post-processing step is applied to merge the insignificant regions. The algorithm is based on mean color, texture and boundary

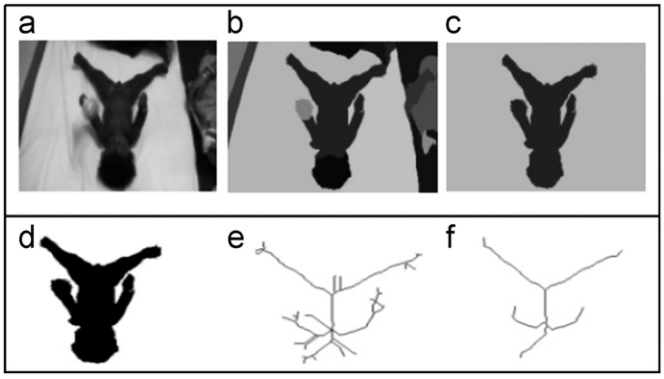

Fig. 2. (a) Original image. (b) Segmented image by mean-shift [13] algorithm with spatial bandwidth $(h_s) = 7$, range bandwidth $(h_r) = 12.5$, and minimum region size 200 pixels. (c) Result after region processing with $T_1 = 0.15$ and $T_2 = 0.4$. (d) Binarized image. (e) Thinned structure using SPTA [24]. (f) Result after applying the filtering algorithm considering $T_3 = 0.1$. The images were recorded during the HINE adductors angle measurement test in the neuro-development clinic of Institute of Post Graduate Medical Education & Research (IPGMER) and Seth Sukhlal Karnani Memorial (SSKM) Hospital, Kolkata, India.

sharing. The filtering method is capable of merging insignificant regions with relatively important segments. The algorithm estimates the merging likelihoods of the neighboring regions of an insignificant segment. Suppose an image (I) is segmented into N regions. A list of regions (L) is prepared where each node contains the description of a region with parameters like region number, size, boundary length, lengths of shared boundaries with neighboring regions, mean color, and texture details. In case of mean color, averages of Hue (H), Saturation (S), and Intensity (V) are computed. Features described in the work of Tamura et al. [18] are used to extract texture information. It is found that the six textural features proposed in the work of [18] performs satisfactorily when compared with psycho-visual measurements using human subjects. However, the authors have found that out of the six measures, the first three, namely coarseness, contrast, and directionality produce acceptable results for the majority of image classes [18]. Thus, these features are often used in classification. Coarseness (C), Contrast (T) and Directionality (D) of a region are estimated as mentioned in [18]. Finally, a six dimensional feature vector is produced for each region.

$$R_i = [H(i),S(i),V(i),C(i),T(i),D(i)] \tag{1}$$

Suppose the feature vector of a region is represented by R_i. It is defined in (1). A Region Association Graph (RAG) is built. Initially, the RAG contains N nodes representing the segments. A node pair of the graph is linked by an edge if the regions representing the nodes are neighbors. The RAG is used to filter out small and insignificant regions using the algorithm discussed below.

1. Arrange the list of nodes L in ascending order of size and number them as $R_1,R_2,\ldots,R_N$.
2. If $|Size(R_1)| \le T_1 \times |Size(R_N)|$, go to step 3, else go to step 7. T_1 is a threshold that is estimated experimentally. Since the backgrounds of the images being considered are expected to be larger than the foreground objects (segments representing the baby being tested), T_1 can be estimated using average foreground/background ratio. It has been found that with fixed camera position and zooming, foreground objects cover on an average 15–20% of the pixels of the images. Thus, a value of $T_1 = 0.15$ works reasonably well for the images of HINE adductors angle measurement.
3. An adjacency list is created using the neighbors of R_1 subject to the condition that the neighbors satisfy (2) where the distance between the two six dimensional vectors is computed using L2 norm. T_2 is a threshold that is used to check whether two neighboring regions are having similar color and texture or not. The foreground object is expected to be dominated by skin pixels and the background is largely white. Thus, a higher value of T_2 can be assumed. Experimentally is has been found that a value of $T_2 = 0.4$ works satisfactorily for the class of images being processed.

$$|R_1 - R_{neighbor}| \le T_2 \tag{2}$$

4. *Initialization*: Suppose there are k neighbors of R_1 that satisfy (2). Initially, it is assumed that all of the k neighbors are equiprobable to be merged with R_1 subject to the condition mentioned in (3).

$$\sum_{l=1}^{k} P_{merge}(R_1,R_1^l) = 1 \tag{3}$$

5. *Likelihood estimation*: Merging likelihoods of the neighbors are updated using (4) where $B(x)$ denotes the boundary of region x. Since, seven components are used in the merging likelihood estimation, the computed value of the likelihood is divided by seven to give equal importance to all the components.

$$P_{merge}(R_1,R_1^l) = \frac{1}{7}\left[\frac{B(R_1)\cap B(R_1^l)}{\sum_{l=1}^{k} B(R_1)\cap B(R_1^l)} + \sum_{i=1}^{6} \frac{|R_1(i)-R_1^l(i)|}{\sum_{l=1}^{k} |R_1(i)-R_1^l(i)|}\right] \tag{4}$$

6. *Update*: Sort the adjacency list of R_1 in ascending order based on the probability values estimated in step 5. Select the first node of the list, say y, and merge R_1 with y and update the node representing y using (5) and (6). Information regarding boundary sharing of y with the newly formed neighboring regions is also updated. Recalculate the mean color and texture information of y and go to step 1.

$$|size(y)| = |size(R_1)| + |size(y)| \tag{5}$$

$$|B(y)| = |B(R_1)| + |B(y)| - |B(R_1)\cap B(y)| \tag{6}$$

7. The final set of segments is generated where every segment is expected to be significant.

To explain the region processing scheme in detail, the synthetic image as shown in Fig. 3(a) is used. Let, the image be partitioned into seven non-overlapping segments/regions. Assume that R_7 is the smallest segment and it has boundary sharing with other segments, e.g., R_1, R_4, R_5, and R_6. Suppose, the merging likelihood of R_6 is estimated to be the highest amongst all the neighboring regions of R_7. In such a situation, R_7 should be merged with R_6 and corresponding nodes should be updated. In Fig. 3(b)–(d), images after region merging, initial RAG, and updated RAG are shown. The process continues till the condition mentioned in step 2 is valid. The proposed algorithm has been applied on the outputs of mean-shift based segmentation to merge insignificant regions with nearby important regions. It is found that the algorithm performs satisfactorily for the images being considered in this paper. An example of region merging is depicted in Fig. 2(c). The segmentation of the image shown in Fig. 2(b) is processed with the proposed algorithm. The values of T_1 and T_2 were set to 0.15 and 0.4 respectively.

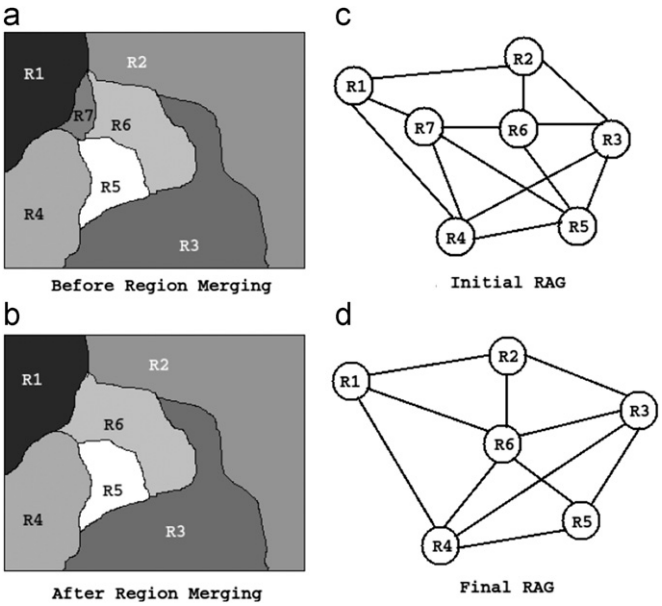

Fig. 3. (a) A synthetic segmentation. (b) Output after region merging using the proposed algorithm. (c) Initial Region Association Graph (RAG). (d) Updated Region Association Graph (RAG).

2.3. Thinning and filtering

A method to separate the foreground object containing the infant undergoing neurological examinations is discussed in this part. Since the baby is unclothed before the examination commences and the background is kept white, skin color based foreground separation is possible. The association graph generated using the region merging algorithm discussed in the preceding section can be used for foreground detection. Since the mean color of a region is stored in the data structure of the node representing the region, it is used for generating foreground statistics. There exist a number of algorithms for detecting skin pixels based on statistical information [19–21]. These algorithms are often used in human face and skin region segmentation in image as well as video based applications. It has been reported in the literature that probabilistic estimation of skin pixels is possible if perceptual color spaces like HSV, HSL, and TSL [22] are used. However, a good skin classifier should be able to discriminate between skin and non-skin pixels for a wide range of samples with different skin types such as white, yellow, brown and dark. The classifier should be able to perform well under different illumination conditions such as indoor, outdoor and with white and non-white illumination sources.

Since the illumination level is controlled during video recording of HINE process, the histogram based model proposed by Jones et al. has been found to be effective [22]. It is computationally efficient than the parametric models such as Mixture of Gaussian [23]. After foreground object identification, a binary image is produced such that the black pixels represent foreground (expected to be infant object) and white pixels represent background. Thinning of binary objects is performed next. It has been noted that the scheme based on Safe Point Thinning Algorithm (SPTA) proposed by Naccache et al. is capable of producing good results for varying binary structures [24]. The algorithm iteratively removes the boundary points of the object from all directions. It is illustrated with an example shown in Fig. 4. SPTA based thinning algorithms perform satisfactorily on varying shaped objects. Since the body parts of an infant are expected to form shapes like the one shown in Fig. 4(b), SPTA produces thinned structures which are noisy in the context of current application. This leads to wrong estimation of features.

As a remedy, the heuristic based filtering proposed in [12] has been used to remove insignificant edges. The approach begins with the SPTA based thinning algorithm. Let, $G(V,E)$ be the graph of the thinned structure. In the graph, the line segments of the skeleton correspond to edges and joining or end points of the segments represent the nodes. An edge is assumed to be a hanging edge when one of its vertices is a leaf node. The set of hanging edges is denoted by H where $H \subseteq E$. Next, a greedy algorithm is adopted that iteratively removes the smallest edge out of the edges hanging from an intermediate vertex. The refinement process is described in Algorithm 1.

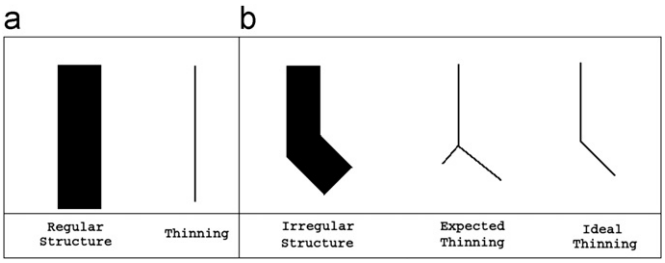

Fig. 4. Synthetic examples of thinning and its limitations to produce desired results when used in thinning structures that are often found after binarizing the images of infants being tested for HINE.

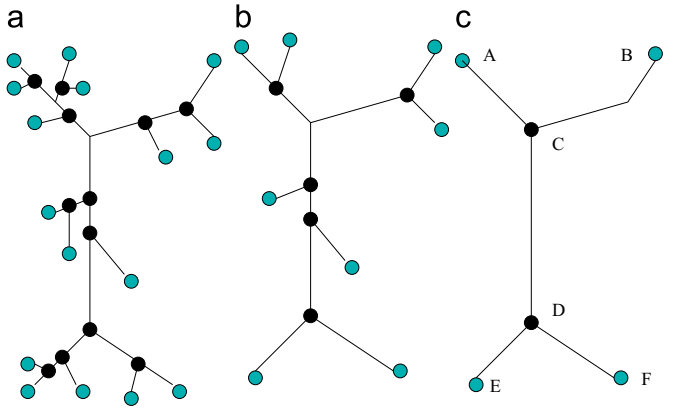

Fig. 5. Results obtained at various intermediate steps of the filtering process using a synthetic example. (a) Initial synthetic thinned image. (b) Intermediate thinned image. (c) Final thinned image.

Algorithm 1. Filtering of thinned object after SPTA based skeletonization.

Input: Graph representing the thinned image $G(V,E)$ and H
Output: Graph representing filtered/refined image
forall the $h \in H$ **do**
  if *length* $(h) < T3 \times$ *size of the longest edge* **then**
    $H = H - h$;
    $E = E - h$;
forall the $h_1 \in H$ *AND* $h_2 \in H$ **do**
  If h_1 *AND* h_2 *are connected to same intermediate vertex* **then**
    If *length* $(h_1) >$ *length*(h_2) **then**
      $E = E - h_1$;
    else
      $E = E - h_2$;
If *No change in the graph in successive iterations* **then**
  Stop the algorithm and **return**$\{G(V,E)\}$;

The algorithm iteratively removes small and insignificant hanging edges. Finally, it produces a filtered graph containing only the significant edges. Intermediate steps of the proposed filtering/refinement scheme are elaborated with a synthetic example in Fig. 5. Internal and leaf nodes are highlighted with separate colors for clarity. After carrying out several experiments with the images of HINE adductors angle measurement, it has been found that the noisy edges can be removed if the value of T_3 is set to 0.1. Output of the proposed filtering scheme is presented in Fig. 2. It is evident that the filtered image is less noisy than the original thinned image. In majority of the cases, insignificant edges are removed. Using these filtered structures, features can be extracted to measure adductors angle.

3. Analysis using single image

The filtering scheme described in Algorithm 1 is applied to estimate adductors angle. Assume that the graph shown in Fig. 5(c) represents the thinned and filtered image of the baby undergoing this test. Suppose the junction and end points of the graph are labelled as mentioned in the figure. The camera is mounted in such a way that the scene can be recorded from the top. The infant is placed in north–south orientation as shown in Fig. 1. Such an arrangement ensures that the joining point between two legs at supine posture is easily identified with geometrical information of the filtered skeleton

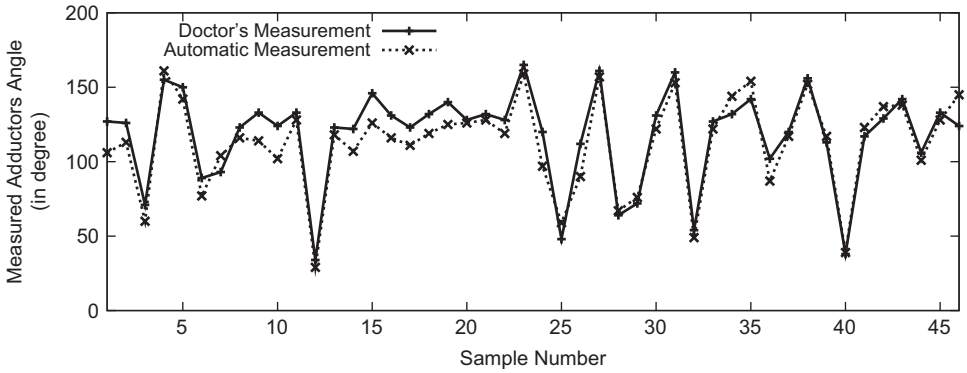

Fig. 6. Comparison of automatic adductors angle measurement with respect to ground truths recorded in the neuro-development clinic of Institute of Post Graduate Medical Education & Research (IPGMER) and Seth Sukhlal Karnani Memorial (SSKM) Hospital, Kolkata, India.

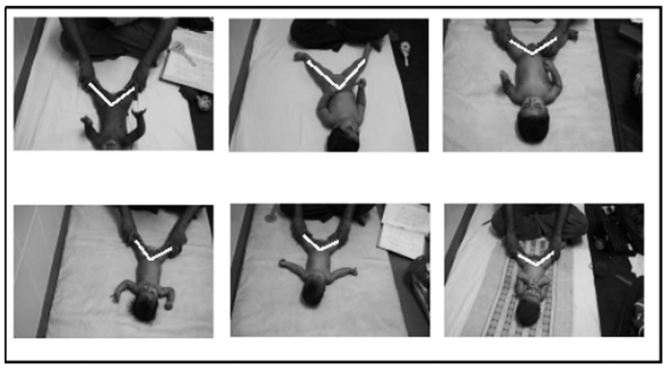

Fig. 7. Results of abductors angle measurement using some of the sample images collected from IPGMER and SSKM Hospital.

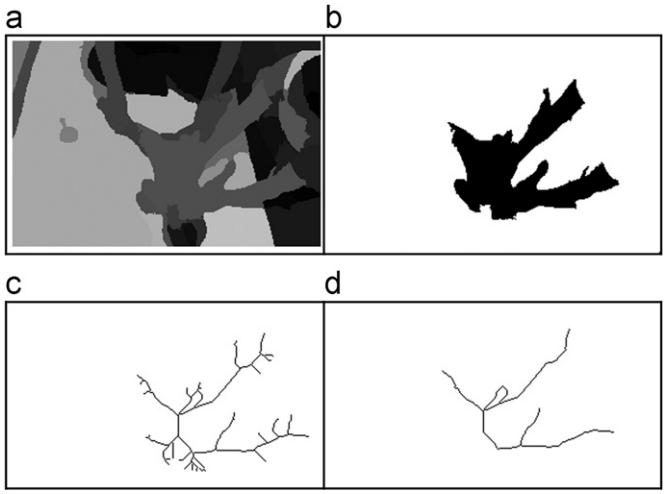

Fig. 8. Demonstration of errors at various steps of feature extraction process used for the analysis of adductors angle measurement examination. (a) Segmented image. (b) Binarized image. (c) Skeleton image. (d) Filtered skeleton image.

image. The junction point C is detected and two straight lines are fitted along the edges maneuvered over the legs of the baby using least square approximation method. Finally, the angle between the line segments $\overline{AB}$ and $\overline{AC}$ as mentioned in Fig. 5(c) is estimated.

Images and videos of adductors angle measurement for 50 infants were recorded at the neuro-development clinic of SSKM Hospital, Kolkata, India. The proposed filtering algorithm was capable of generating usable skeletons for 46 images. Comparative results using the proposed scheme with respect to ground truths are shown in Fig. 6. Ground truths were recorded by the experts. In the graph, x-axis represents sample number and y-axis represents measured angle (in degree). It is evident from the graph that the results agree with ground truth measurements (angle measured by doctors) in majority of the cases. Above 90% precision and recall have been recorded. Some of the results of adductors angle measurement are depicted in Fig. 7. Two thick straight lines (painted white) along the legs of the baby are shown in the figure for visualization.

3.1. Failed cases

The effect of errors in segmentation, thinning, and filtering lead to erroneous results in some cases. In four of the sample cases, feature extraction was unable to produce usable features from the images. Thus, the angle measurement could not be completed using the heuristic described earlier for these images. An example of erroneous thinning and filtering is shown in Fig. 8. It is evident that due to complicated skeleton image, the junction point between the legs of the infant being tested may not be identified unambiguously. In the next section, a video based scheme is discussed to address such erroneous conditions.

4. Video based analysis

The automation process can be improved by analyzing videos of the examination. It is based on feature point tracking. In this work, a semi-automatic approach is adopted. During offline analysis, the examiner was asked to mark locations of left knee, right knee, and junction between the legs of the baby in the first frame of the video. The marked points are tracked successively in the temporal domain. However, selecting a suitable tracking algorithm for the application under consideration is important. Some of the popular tracking algorithms are human blob tracking [25], kernel based tracking [26], target tracking [27] and feature point based tracking [28]. Majority of these algorithms rely upon matching between successive frames in block or macro-block levels. Though block or macro-block matching schemes are primarily used in video compression, often they are used in designing tracking algorithms. Some of the examples of block matching algorithm are Full Search Block Matching (FSBM) [29], Diamond Search (DS) [30,31], Directional Diamond Search (DDS) [32], Three Step Search (TSS) [33], and Efficient Three Step Search (ETSS) [34].

These schemes help in searching the probable location of a block of the current frame in the next frame. However, block matching based schemes are inefficient in real time tracking due to their high computational complexity. Thus, instead of detecting the motion of an entire object, selected feature points of the object can be tracked. Feature points are defined as locations on

the object with distinct appearance or characteristics. Estimating feature point based motion is computationally efficient. Thus, object tracking schemes often adopt the idea of feature point based motion estimation. Some of the popular feature point based object tracking schemes can be found in [35–37]. In this work, a dynamic programming and node pruning based feature point tracking algorithm has been used [10]. The algorithm is computationally efficient and works satisfactorily for the class of videos being considered in this paper.

The tracking can be applied as follows. Assume that the feature points are marked in the first frame. Since the physician stretches the legs of the infant during this examination, it is expected that the angle between the legs will increase continuously as long as stretching is continued. This can be measured from the tracked feature points. However, once stretching is stopped due to the resistance experienced by the physician, the value of the angle does not change anymore and the measurement is recorded. In Fig. 9, variation of angle in temporal domain is shown for 23 sample videos. It may be observed that no significant change in the angle occurs when stretching is stopped. The variation of angle can be approximated using the curve defined in (7). A graphical representation of the curve is shown in Fig. 10. If the value of $(a+b)$ is estimated, a decision regarding the saturated value of the adductors angle may be taken. Thus, a regression analysis is performed such that the values of a, b, and c can be estimated.

$$y = a + b(1-e^{-cx}) \tag{7}$$

The approximation is done as follows. Let, the squared error in estimation given in (8) be denoted by E. To minimize the error, first order partial differentiations of the error with respect to a, b, and c are computed using (9)–(11). The optimal values of a, b, and c are obtained by iterative modification of their values as in (12)–(14).

$$E = [y-a-b(1-e^{-cx})]^2 \tag{8}$$

$$\frac{\partial E}{\partial a} = -2[y-a-b(1-e^{-cx})] \tag{9}$$

$$\frac{\partial E}{\partial b} = -2[y-a-b(1-e^{-cx})][1-e^{-cx}] \tag{10}$$

$$\frac{\partial E}{\partial c} = -2[y-a-b(1-e^{-cx})][bxe^{-cx}] \tag{11}$$

Algorithm 2 begins with the initial value of y (say y_0). It is computed using the initial values of a_0, b_0, and c_0 which have been set to 0.01, 0.01 and 0.2 respectively for this experiment. The algorithm iterates over n observed values $(y_0, y_1, \ldots, y_{n-1})$. It checks for convergence with a threshold T using a learning rate γ. The values of T and γ are set to 10^{-5} and 5×10^{-4} respectively. The algorithm converges quickly even with small values of γ and T as on an average it iterates for 100 sample points (the average number of frames in a sample video recorded at 25 fps is 100).

$$a_{i+1} = a_i - \gamma\frac{\partial E}{\partial a_i} \tag{12}$$

$$b_{i+1} = b_i - \gamma\frac{\partial E}{\partial b_i} \tag{13}$$

$$c_{i+1} = c_i - \gamma\frac{\partial E}{\partial c_i} \tag{14}$$

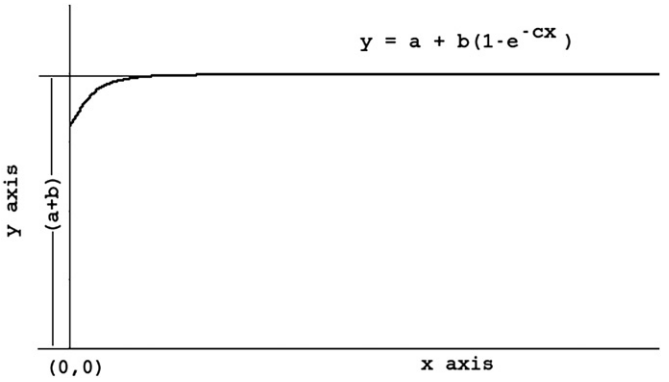

Fig. 10. An example of the exponential curve used to fit the data where x-axis represents the frame number and y-axis represents the measured angle in degree.

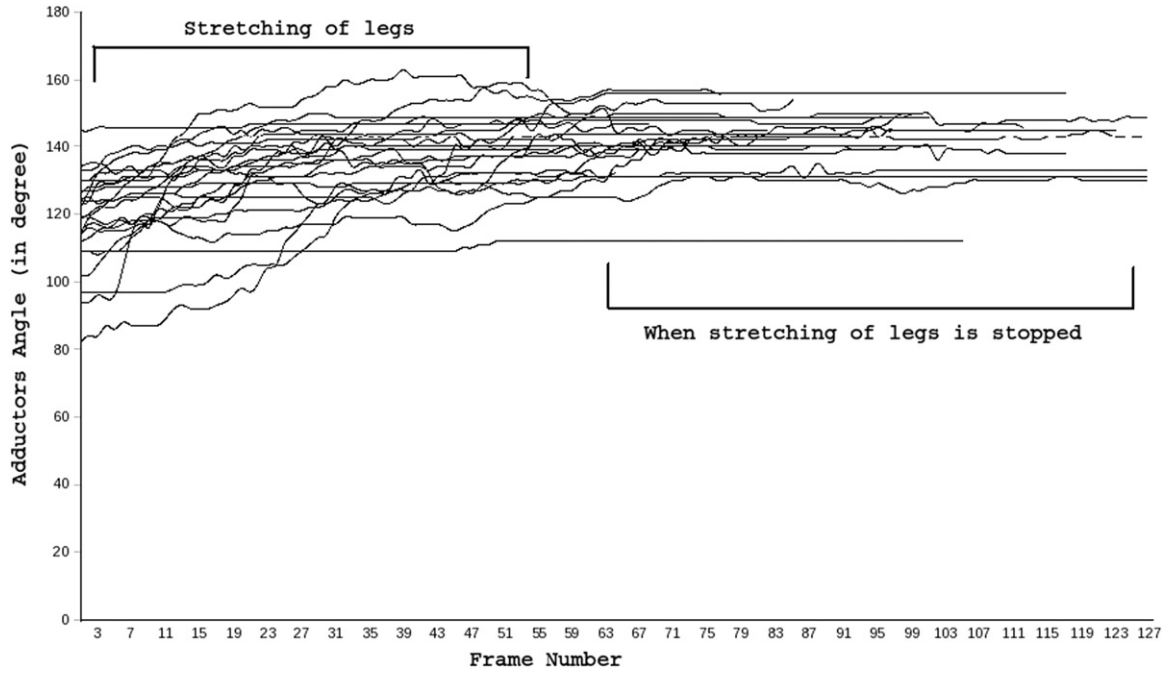

Fig. 9. Variation of angle during adductors angle measurement using feature point based object tracking.

Algorithm 2. Iterative computation of the values of a, b, and c.

Input: Learning rate: γ, Number of points: n, Convergence threshold: T
Output: Computed values of a, b, and c
1 Initialize $i=0$, $a_0=0.01$, $b_0=0.01$, and $c_0=0.2$;
2 Estimate a_1, b_1, and c_1 using (12)–(14) at $y=y_0$;
3 **while** $[(a_{i+1}-a_i)^2+(b_{i+1}-b_i)^2+(c_{i+1}-c_i)^2]<T$ *or* $i>n-1$ **do**
4 $i=i+1$;
5 Compute $\frac{\partial E}{\partial a_i}$, $\frac{\partial E}{\partial b_i}$ and $\frac{\partial E}{\partial c_i}$ *at* $y=y_i$ using (9)−(11);
6 Update a_{i+1}, b_{i+1}, and c_{i+1} using (12)−(14);
7 **return** a_i, b_i, and c_i;

In Fig. 12, graphical representation of the measured adductors angle variation (Sample 21) and its approximation using exponential regression are shown. The values of a, b, and c are found to be 77.64, 77.64, and 0.15 respectively. Based on these estimated values, the adductors angle is found to be 155.28° which is very close to the ground truth measurement (154°). Some of the video frames of "Sample 21" are shown in Fig. 11. It may be observed that the angle painted using thick white straight lines is increasing with time and remains almost constant beyond Frame 60. In Table 1, estimated values of a, b, and c for the 23 samples shown in Fig. 9 are presented. Corresponding ground truths suggest that the video based scheme accurately estimates the adductors angle in majority of the cases.

It is possible to develop a fully automated scheme with the help of image based analysis discussed in Section 3. Such a method can be used for the detection of feature points in the first frame of the video and the points can be tracked in successive frames. However, it may not be possible to get desirable skeleton images in all cases. Any error in estimating the three locations for measuring adductors angle in the first frame will make the tracking erroneous. Thus, it may not be wise to use feature point detection followed by tracking in this context. However, successive video frames can be processed individually to extract features and then a temporal correspondence can be obtained. The effect of the errors in segmentation and thinning can be minimized using such an approach. It is unlikely that all frames of a video would be affected by segmentation and thinning errors. A correct decision is possible by virtue of majority voting or consensus agreement even if some of the frames get affected by such errors. This has been adopted to make the analysis fully automatic.

It has been observed that the adductors angle estimation process lasts for 4–5 s. Thus, 100–125 frames need to be processed if the videos are recorded at 25 fps. If a desktop computer powered by Intel Dual Core processor @2.13 GHz, 3 GB of DDR2 memory is used, mean-shift based segmentation [13] usually takes approximately 4 s to segment an image of size 320×240 with spatial bandwidth

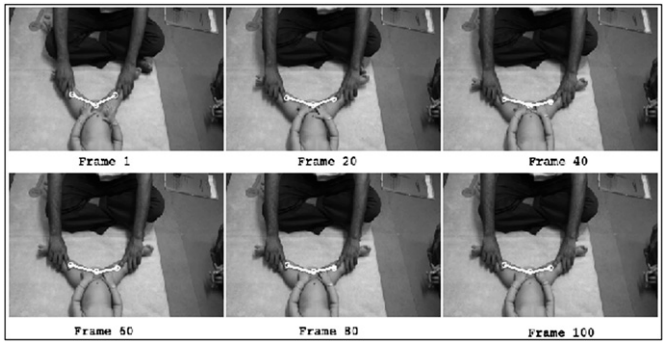

Fig. 11. Detection of adductors angle using semi-automatic approach using video based analysis. The images have been extracted from the videos of Sample 21 of the dataset.

Table 1
Estimated values of a, b, and c using the proposed curve fitting algorithm applied on 23 sample videos of adductors angle measurement.

Sample number	Number of frames	Value of a	Value of b	Value of c	Value of $a+b$	Ground truth
$T=0.000001$, $\gamma=0.0005$, $a_0=0.01$, $b_0=0.01$, $c_0=0.2$						
1	127	64.65	64.65	0.14	129	124
2	100	74.49	74.49	0.17	149	146
3	127	65.39	65.39	0.14	131	139
4	105	55.89	55.89	0.16	112	100
5	99	75.88	75.88	0.17	152	149
6	98	69.89	69.89	0.17	140	134
7	133	66.39	66.39	0.14	133	142
8	79	77.89	77.89	0.21	156	152
9	90	72.15	72.15	0.19	144	131
10	68	73.15	73.15	0.24	146	139
11	47	74.15	74.15	0.33	148	144
12	117	68.96	68.96	0.15	138	137
13	64	65.98	65.96	0.25	132	131
14	126	71.46	71.46	0.14	143	150
15	82	72.47	72.46	0.20	145	146
16	123	72.46	72.46	0.14	145	142
17	85	76.37	76.36	0.20	153	153
18	103	69.96	69.96	0.16	140	138
19	111	55.13	55.16	0.59	110	112
20	127	74.15	74.15	0.139	148	146
21	54	77.65	77.65	0.15	155	154
22	68	69.15	69.15	0.24	138	129
23	101	74.15	74.15	0.17	148	141

$(h_s)=7$, range bandwidth $(h_r)=12.5$, and minimum region size= 500 pixels. Thus, if mean-shift is applied to all the frames of a video, it takes more than 7–8 min to process one complete video. This is a time consuming process. To overcome this, only the *I-frames* or intra-coded frames of a video are used for feature extraction. *I-frame* is the first frame of a group of pictures (termed as GOP in video encoder/decoder literature. GOP is usually set to a value of less than 16). An *I-frame* contains the information of the image in intra-prediction mode, whereas other frames of a GOP are predicted. However, to extract these frames, a video decoder is necessary. Intel OpenCV library [38], which supports frame extraction has been used in our implementation. Mean-shift segmentation, thinning and filtering takes approximately 5 s to process an image of dimension 320×240 with the hardware configuration discussed above. *I-frame* based scheme takes approximately 40–50 s (8–10 *I-frames*) to process a video of 5 s duration. This is significantly low as compared to the processing time using all of the frames (7–8 min to process a single video). Thus, in this analysis, *I-frame* based scheme is preferred.

In Fig. 13, results of segmentation and thinning applied on the *I-frames* of the video of "Sample 19" are shown. It may be observed that due to errors in segmentation and thinning, some of the frames, e.g., 5, 10, 50, 95, 110, and 115 are not useful for measuring the angle. However, angle can be measured from the remaining frames. A graphical representation of the measured angle corresponding to the outputs is shown in Fig. 14. An approximation of the data points using the exponential curve described in (7) is also shown in the figure. The coefficients of the exponential curve have been found to be $a=55.16$, $b=55.13$, and $c=0.59$. Using these parameters, the approximated adductors angle for the baby undergoing this examination has been estimated to be 110.29°. This is closely matching with the ground truth (112°) measured by the physician using Goniometer.

5. Comparative results

In this part of the discussion, a comparative performance analysis of the three schemes proposed in this paper is presented.

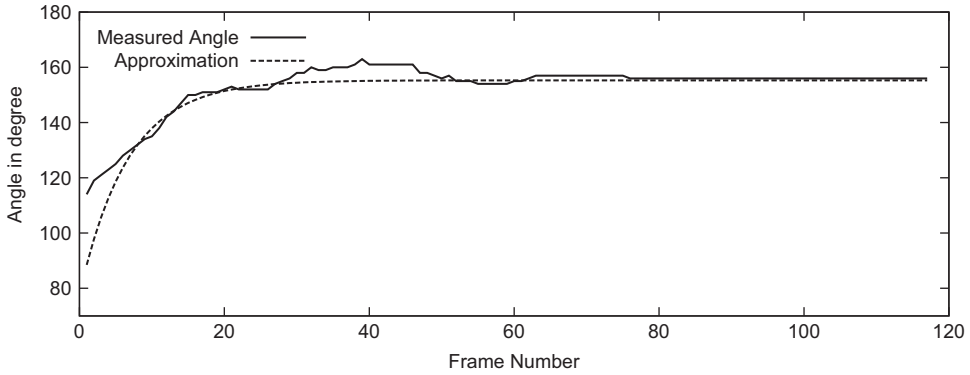

Fig. 12. Variation of adductors angle (Sample 21) measured using video based analysis and its approximation using the exponential regression analysis.

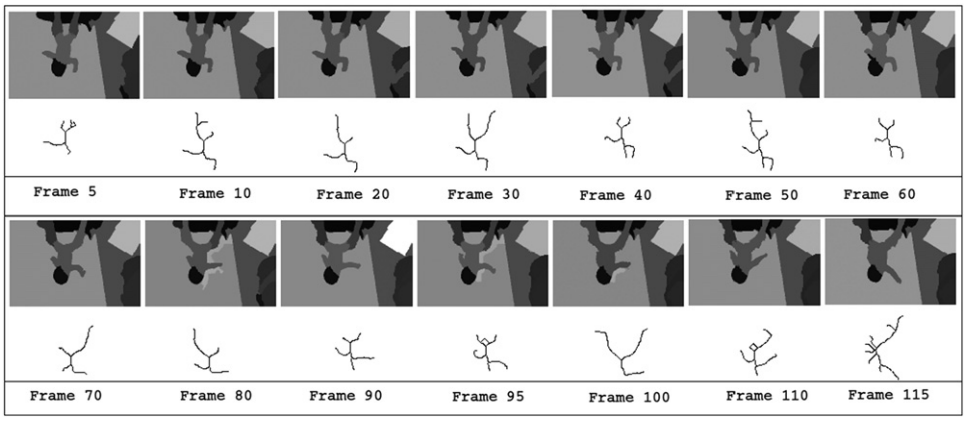

Fig. 13. Segmentation and thinning steps applied on the *I-frames* of the video of Sample 19.

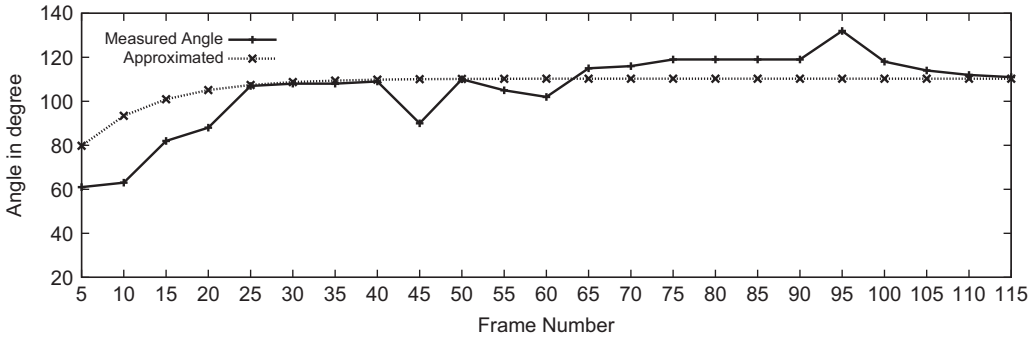

Fig. 14. Graphical representation of the varying adductors angle for the segmented and skeletonized images shown in Fig. 13 and its approximation using exponential curve fitting.

HINE adductors examination is graded based on the ranges of angles and a score between 0 and 3 is awarded. According to the recommendation given in [4], a score of 3 signifies that the tone of the infant undergoing this test is normal (baby's legs are flexible within the specified range). However, if the physician is unable to stretch the legs beyond 80° (feel the resistance within this angle), the tone of the baby is assumed to be considerably low whereas the other two scores signify that the baby needs further observation and parents are counselled. Since the scores are estimated based on a range of angle, large errors in the measurement process are unlikely. In Table 2, classification results of the single image based scheme with respect to ground truths are presented. Ground truths were recorded by single expert who conducted the examinations at the follow-up clinic of SSKM Hospital during the period of the study. According to the HINE specification, a score of one is not used in adductors assessment. It is found that 46 samples out of 50 are usable for feature extraction. The remaining four images cannot be used due to the errors in segmentation and thinning.

Table 2
Classification of results obtained using single image analysis method discussed in Section 3 with respect to original scores as mentioned in the HINE guidelines.

	$80^\circ < \theta < 150^\circ$ Score: 3			$150^\circ < \theta < 170^\circ$ Score: 2			$\theta < 80^\circ$ Score: 0		
Ground truth (46)	33			6			7		
Outcomes	TP	FP	FN	TP	FP	FN	TP	FP	FN
	32	1	1	6	1	0	7	1	0
Precision	96.96%			85.71%			87.5%		
Recall	96.96%			100%			100%		

TP: True positive, FP: False positive, FN: False negative

Table 3
Classification of results with respect to the original scores as mentioned in the HINE guidelines using the proposed video based analysis scheme.

	$80^\circ < \theta < 150^\circ$ Score: 3			$150^\circ < \theta < 170^\circ$ Score: 2			$\theta < 80^\circ$ Score: 0		
Ground truth (50)	36			7			7		
Outcomes	TP	FP	FN	TP	FP	FN	TP	FP	FN
	35	1	0	6	0	1	7	0	0
Precision	97.22%			100%			100%		
Recall	100%			87.5%			100%		

TP: True positive, FP: False positive, FN: False negative

Table 4
Classification of results with respect to the original scores as mentioned in the HINE guidelines using the proposed image based analysis applied on *I-frames*.

	$80^\circ < \theta < 150^\circ$ Score: 3			$150^\circ < \theta < 170^\circ$ Score: 2			$\theta < 80^\circ$ Score: 0		
Ground truth (50)	36			7			7		
Outcomes	TP	FP	FN	TP	FP	FN	TP	FP	FN
	34	2	0	6	0	1	6	0	1
Precision	94.4%			100%			100%		
Recall	100%			87.5%			87.5%		

TP: True positive, FP: False positive, FN: False negative

Next, in Table 3, classification results with respect to ground truths using the proposed video based analysis are presented. Ground truths were recorded by one physician who carry out the HINE test in the follow-up clinic of SSKM hospital during the period of the study. We have recorded the angle measured using Goniometer by the physician and video recordings were done during the tests. These videos have been analyzed offline and compared with the corresponding ground truths. It may be observed that the algorithm is successful in estimating adductors angle for all samples including the four for which the image based algorithm was unsuccessful. In addition to that, precision values have increased significantly. However, the recall value for *Score*2 has decreased. This is because of the fact that using the video based analysis, value of the adductors angle has been estimated to be 152° when the ground truth suggests that the actual value is 149°. That is why "Sample 5" has been misclassified with *Score*3 instead of *Score*2. This is due to the fact that even a small amount of error such as 1° to 2° can essentially lead to wrong classification when the angle is close to the inter-class boundaries. In Table 4, classification results of HINE adductors angle measurement using the proposed image based analysis applied on *I-frames* are presented. It may be observed that the results have improved when compared to the analysis using single image.

6. Conclusion and future directions

Image and video based algorithms can be used for analyzing infant neurological examinations. In this paper, analysis of HINE adductors angle measurement using image segmentation, feature extraction, and video object tracking is presented. First, it has been discussed that using image segmentation, thinning and examination specific geometrical model the analysis of this examination can be automated. Next, a video based analysis using object tracking has been discussed to improve the accuracy of single image based analysis. The approach is semi-automatic. However, if a fully automatic scheme is required then the image based algorithm is preferable.

A video based automatic analysis can reduce the workload of the physicians involved in conducting the HINE process. The image based algorithm fails in some cases to precisely measure the angle. In such cases, video based analysis may be used. Analysis of other examinations like measuring the popliteal angle, curvature of the torso in ventral suspension and ankle dorsiflexion can be automated using similar techniques.

Conflict of interest statement

None declared.

Acknowledgments

The work has been funded by Ministry of Communication and Information Technology under Approval No. 1(4)/2009-ME& TMD (28-08-2009), Department of Information Technology, Govt. of India. The authors thank Asish Ghosh of the neuro-development clinic of SSKM Hospital for his valuable suggestions related to the HINE process. The work has been carried out by maintaining the ethical practices followed in SSKM Hospital, Kolkata, India. Consents for recording the visuals of the examinations carried out on subjects participated in this study were taken from respective authorities.

References

[1] D. Romeo, A. Guzzetta, M. Scoto, M. Cioni, P. Patusi, D. Mazzone, M. Romeo, Early neurologic assessment in preterm-infants: integration of traditional neurologic examination and observation of general movements, Eur. J. Paediatr. Neurol. 12 (3) (2008) 183–189.
[2] D. Romeo, M. Cioni, M. Scoto, L. Mazzone, F. Palermo, M. Romeo, Neuromotor development in infants with cerebral palsy investigated by the Hammersmith infant neurological examination during the first year of age, Eur. J. Paediatr. Neurol. 12 (1) (2008) 24–31.
[3] B. Vohr, L. Wright, A. Mele, J. Veter, J. Steichen, N. Simon, D. Wilson, S. Broyles, C. Bauer, V. Delaney-Black, K. Yolton, B. Fleisher, L. Papile, M. Kaplan, Neurodevelopmental and functional outcomes of extremely low birth weight infants in the national institute of child health and human development neonatal research network, J. Pediatr. 105 (6) (2000) 1216–1226.
[4] L. Dubowitz, V. Dubowitz, E. Mercuri, The Neurological Assessment of the Preterm and Full Term Infant, vol. 9, Clinics in Developmental Medicine, London, Heinemann, 2000.
[5] S. Singh, H. Hsiao, Infant telemonitoring system, in: Engineering in Medicine and Biology Society, Proceedings of the 25th Annual International Conference of the IEEE, vol. 2, September 2003, pp. 1354–1357.
[6] J. Bhatt, N. Lobo, M. Shah, G. Bebis, Automatic recognition of baby gesture, in: Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence, November 2003, pp. 610–615.
[7] Y. Nishida, Y. Motomura, K. Kitamura, H. Mizoguchi, Infant behavior simulation based on an environmental model and a developmental behavior model, in: Proceedings of IEEE International Conference on Systems, Man and Cybernetics, vol. 2, October 2004, pp. 1555–1560.
[8] D.P. Dogra, K. Nandam, A.K. Majumdar, S. Sural, J. Mukhopadhyay, B. Majumdar, S. Mukherjee, A. Singh, A user friendly implementation for efficiently conducting hammersmith infant neurological examination, in: IEEE International Conference on E-Health Networking, Application and Services, July 2010, pp. 374–378.
[9] D.P. Dogra, K. Nandam, A.K. Majumdar, S. Sural, J. Mukhopadhyay, B. Majumdar, S. Mukherjee, A. Singh, A tool for automatic Hammersmith infant neurological examination, Int. J. E-Health Med. Commun. 2 (2) (2011) 1–13.
[10] D.P. Dogra, A.K. Majumdar, S. Sural, J. Mukherjee, S. Mukherjee, A. Singh, Toward automating Hammersmith pulled-to-sit examination of infants using feature point based video object tracking, IEEE Tran. Neural Syst. Rehabil. Eng. 20 (1) (2012) 38–47.
[11] Goniometer, 〈http://en.wikipedia.org/wiki/Goniometer〉, May 2011.
[12] D. Dogra, A. Majumdar, S. Sural, J. Mukhopadhyay, S. Mukherjee, A. Singh, Automatic adductors angle measurement for neurological assessment of post-neonatal infants during follow up, in: Pattern Recognition and Machine Intelligence, in LNCS Proceedings of the Fourth International Conference, vol. 6744, June 2011, pp. 160–166.
[13] D. Comaniciu, P. Meer, Mean shift: a robust approach toward feature space analysis, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2002) 603–619.

[14] P. Felzenszwalb, D. Huttenlocher, Efficient graph-based image segmentation, Int. J. Comput. Vision 59 (September) (2004) 167–181.
[15] Y. Kuan, C. Kuo, N. Yang, Color-based image salient region segmentation using novel region merging strategy, IEEE Trans. Multimedia 10 (5) (2008) 832–845.
[16] W. Cai, J. Wu, A. Chung, Shape-based image segmentation using normalized cuts, in: IEEE International Conference on Image Processing, October 2006, pp. 1101–1104.
[17] R. Unnikrishnan, C. Pantofaru, M. Hebert, Toward objective evaluation of image segmentation algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 29 (June) (2007) 929–944.
[18] H. Tamura, S. Mori, T. Yamawaki, Textural features corresponding to visual perception, IEEE Trans. Syst. Man Cybern. 8 (1978) 460–472.
[19] P. Kakumanu, S. Makrogiannis, N. Bourbakis, A survey of skin-color modeling and detection methods, Pattern Recognition 40 (3) (2007) 1106–1122.
[20] L. Sigal, S. Sclaroff, V. Athitsos, Skin color-based video segmentation under time-varying illumination, IEEE Trans. Pattern Anal. Mach. Intell. 26 (7) (2004) 862–877.
[21] V. Vezhnevets, V. Sazonov, A. Andreeva, A survey on pixel-based skin color detection techniques, in: Proceedings of the International Conference on Computer Graphics (GRAPHICON), September 2003, pp. 85–92.
[22] M. Jones, J. Rehg, Statistical color models with application to skin detection, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, June 1999, pp. 274–280.
[23] M. Yang, N. Ahuja, Gaussian mixture model for human skin color and its applications in image and video databases, in: Proceedings of SPIE its Application in Image and Video Databases, 1999, pp. 458–466.
[24] N. Naccache, R. Shinghal, SPTA: a proposed algorithm for thinning binary patterns, IEEE Trans. Syst. Man Cybern. 14 (3) (1984) 409–418.
[25] M. Lee, R. Nevatia, Human pose tracking in monocular sequence using multilevel structured models, IEEE Trans. Pattern Anal. Mach. Intell. 31 (1) (2009) 27–38.
[26] D. Comaniciu, V. Ramesh, P. Meer, Kernel-based object tracking, IEEE Trans. Pattern Anal. Mach. Intell. 25 (5) (2003) 564–577.
[27] H. Song, M. Shen, Target tracking algorithm based on optical flow method using corner detection, Multimedia Tools Appl. 52 (1) (2011) 121–131.
[28] T. Gao, G. Li, S. Lian, J. Zhang, Tracking video objects with feature points based particle filtering, Multimedia Tools Appl. (2011) 1–21.
[29] Y. Lin, S. Tai, Fast full-search block-matching algorithm for motion-compensated video compression, IEEE Trans. Commun. 45 (5) (1997) 527–531.
[30] J. Tham, S. Ranganath, M. Ranganath, A. Kassim, A novel unrestricted center-biased diamond search algorithm for block motion estimation, IEEE Trans. Circuits Syst. Video Technol. 8 (4) (1998) 369–377.
[31] S. Tai, Y. Chen, Y. Chen, Small-diamond-based search algorithm for fast block motion estimation, Image Commun. 22 (November) (2007) 877–890.
[32] H. Jia, L. Zhang, Directional diamond search pattern for fast block motion estimation, Electron. Lett. 39 (22) (2003) 1581–1583.
[33] W. Booth, J. Noras, D. Xu, A novel fast three-step search algorithm for block-matching motion estimation, in: Asian Conference on Computer Vision, vol. 1352, 1997, pp. 623–630.
[34] X. Jing, L. Chau, An efficient three-step search algorithm for block motion estimation, IEEE Trans. Multimedia 6 (3) (2004) 435–438.
[35] B. Han, W. Roberts, D. Wu, J. Li, Robust feature-based object tracking, in: Society of Photo-Optical Instrumentation Engineers Conference Series, vol. 6568, April 2007.
[36] Q. Miao, G. Wang, X. Lin, Y. Wang, C. Shi, C. Liao, Scale and rotation invariant feature-based object tracking via modified on-line boosting, in: IEEE International Conference on Image Processing, September 2010, pp. 3929–3932.
[37] V. Pallavi, J. Mukherjee, A. Majumdar, S. Sural, Ball detection from broadcast soccer videos using static and dynamic features, J. Visual Commun. Image Representation 19 (7) (2008) 426–436.
[38] Intel Corporation, Intel Open Source Computer Vision Library (OpenCV), ⟨http://sourceforge.net/projects/opencvlibrary/⟩, May 2011.