Toward Automating Hammersmith Pulled-To-Sit Examination of Infants Using Feature Point Based Video Object Tracking

Debi P. Dogra, *Student Member, IEEE*, Arun K. Majumdar, *Senior Member, IEEE*, Shamik Sural, *Senior Member, IEEE*, Jayanta Mukherjee, *Senior Member, IEEE*, Suchandra Mukherjee, and Arun Singh

***Abstract*—Hammersmith Infant Neurological Examination (HINE) is a set of tests used for grading neurological development of infants on a scale of 0 to 3. These tests help in assessing neurophysiological development of babies, especially preterm infants who are born before (the fetus reaches) the gestational age of 36 weeks. Such tests are often conducted in the follow-up clinics of hospitals for grading infants with suspected disabilities. Assessment based on HINE depends on the expertise of the physicians involved in conducting the examinations. It has been noted that some of these tests, especially pulled-to-sit and lateral tilting, are difficult to assess solely based on visual observation. For example, during the pulled-to-sit examination, the examiner needs to observe the relative movement of the head with respect to torso while pulling the infant by holding wrists. The examiner may find it difficult to follow the head movement from the coronal view. Video object tracking based automatic or semi-automatic analysis can be helpful in this case. In this paper, we present a video based method to automate the analysis of pulled-to-sit examination. In this context, a dynamic programming and node pruning based efficient video object tracking algorithm has been proposed. Pulled-to-sit event detection is handled by the proposed tracking algorithm that uses a 2-D geometric model of the scene. The algorithm has been tested with normal as well as marker based videos of the examination recorded at the neuro-development clinic of the SSKM Hospital, Kolkata, India. It is found that the proposed algorithm is capable of estimating the pulled-to-sit score with sensitivity (80%–92%) and specificity (89%–96%).**

***Index Terms*—Examination automation, infant neurological examination, pulled-to-sit, video object tracking.**

I. Introduction

ADVANCED neonatal care performed in neonatal intensive care units (NICU) has increased the survival rate of very low birth weight (VLBW) and preterm newborns [10]. However, a significant number of infants born at the gestational age[1] below 34 weeks usually show neurological and developmental disabilities [10]. Also, due to the high risk of intra-ventricular haemorrhage (IVH) or peri-ventricular leukomalacia (PVL), increasing prevalence of Cerebral Palsy (CP) may occur in premature, low birth-weight newborns and children born with asphyxia [13]. In such cases, early prediction using the outcome of various clinical examinations performed in NICU is important to counsel families who substantially benefit from early interventions. It has also been reported in literature that an early diagnosis can reduce the rate of CP [13].

Hammersmith Infant Neurological Examination (HINE) [7] is one such approach to quantitatively assess the neurological development of infants between 2 and 24 months of corrected age[2]. Its purpose is to asses the neuro-motor development of infants [10], [17], [18], [20]. According to the recommendation of Dubowitz *et al.* [7], the neurological development scores can be estimated using the outcomes of neurophysiological examinations that are broadly classified into three categories, namely *neurological items, motor milestones* and *behavior*. Out of these three categories, the first group of tests deals with the neurological development of a baby and it has five subgroups, namely cranial nerve function, posture, movement, tone, and reflex/reaction. Postures and reactions of the infant under consideration are recorded while these examinations are being carried out. All of the tests in this category are graded with scores between 0 to 3 where a score of 3 signifies that the development of the baby is normal and 0 means abnormal. The remaining two groups deal with motor and behavioral development. There is no such standard scoring technique available for motor development whereas assessment of behavioral development is a subjective process.

A. Advantages of Automatic Analysis of HINE

In most clinics, assessment of the HINE is usually done by visual observation. However, experts often face difficulty in grading outcomes of some of the examinations solely based on visual observations. Especially, assessment of movement related examinations such as pulled-to-sit is often difficult. A typical sequence of images of pulled-to-sit examination is shown in Fig. 1. These images were captured at the neuro-development clinic of the Seth Sukhlal Karnani Memorial (SSKM)

[1]Gestational age is the age of an embryo or fetus [11].

[2]Corrected age = chronological age or gestational age – number of weeks or months premature [1].

Manuscript received March 15, 2011; revised August 09, 2011; accepted September 28, 2011. Date of publication December 07, 2011; date of current version January 25, 2012. This work was supported in part by the Ministry of Communication and Information Technology under Approval 1(4)/2009-ME&TMD (28-08-2009) and by the Department of Information Technology, Government of India.

D. P. Dogra, A. K. Majumdar, and J. Mukherjee are with the Department of Computer Science and Engineering, Indian Institute of Technology Kharagpur, 721302 India (e-mail: dpdogra@cse.iitkgp.ernet.in; akmj@cse.iitkgp.ernet.in; jay@cse.iitkgp.ernet.in).

S. Sural is with the School of Information Technology, Indian Institute of Technology Kharagpur, 721302 India (e-mail: shamik@sit.iitkgp.ernet.in).

S. Mukherjee and A. Singh are with the Department of Neonatology, Institute of Post Graduate Medical Education and Research and Seth Sukhlal Karnani Memorial Hospital, Kolkata 700020 India (e-mail: drsmukherjee70@gmail.com; drarunsingh61@yahoo.co.in).

Digital Object Identifier 10.1109/TNSRE.2011.2172223

Hospital, Kolkata, India [3]. It is necessary to quantify relative head movement of the baby with respect to torso for assessment of pulled-to-sit examination. Thus, when the examination is observed in sagittal view (side view), it is expected that a normal baby would lift the head along with the torso. However, an abnormal baby with probable neurophysiological disorder is unlikely to lift the head along with torso. In such cases, a head lag is expected. Experts who conduct HINE at speciality hospitals have pointed out that grading of pulled-to-sit examination becomes difficult when the baby moves other body parts during this examination. In such cases, video-based analysis can lead to a useful decision support tool. In the following part, we list some of the advantages of video-based analysis of such examinations.

1) Since the grading of the examination is done based on the angle between head and torso, the measurement can be affected by observer's perception. There is a chance of interobserver variance while detecting the angle during pulled-to-sit process. This can be minimized using camera-based system. Though the automatic measurement of angle using computer-based system may suffer from perspective problem, it can be avoided using fixed arrangement of camera and examination bed.
2) Unlike a conventional approach where the same test is repeated when the examiner misses some of the key reactions during the test, video-based analysis of the recordings of the pulled-to-sit examination can be done repeatedly with the stored video shot to detect such events.
3) It has been observed that the time required to complete the entire set of examinations on a single baby is approximately 30 minutes in the current setup of the neuro-development clinic of the SSKM Hospital, India. This is due to the fact that infants often do not cooperate with the examiner during examinations like pulled-to-sit, lateral tilting, adductors, popliteal, ankle dorsiflexion, etc. Also, placing the Goniometer [2] across various body parts of the baby for measuring angle is time consuming and error prone.[3] If videos of the examinations are recorded, they can be processed offline and the process can be expedited. This will also reduce the chance of infection due to contactless observation.
4) Video-based automatic or semi-automatic system requires less clinical expertise. Even trained paramedical persons can perform such examinations.
5) HINEs are repeated at various corrected ages like 3, 6, 9, 12 months and onwards. Correlation among the scores estimated at various ages is useful in longitudinal study. Thus, if videos of the examinations conducted at earlier stages are available, it will essentially help the doctors in understanding the nature of neurological development of the baby.

A number of applications in medical informatics domain adopt computer vision-based solutions. For example, Bhatt *et al.* [4] have proposed a video-based infant monitoring tool that successfully recognizes the event when an unattended baby puts something into its mouth. Similarly, the tool developed by Singh *et al.* [16] is a telemonitoring system which can be used for remote surveillance of infants. It was developed using safe, compact and noninvasive sensors to record the movements of a baby

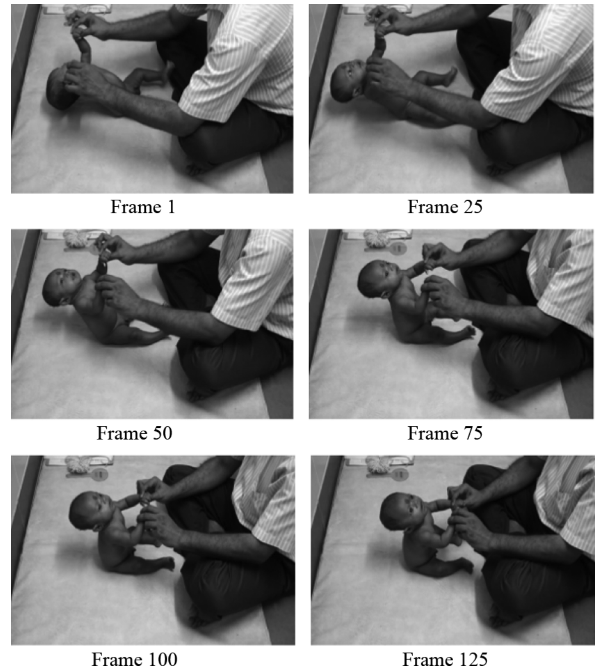

Fig. 1. Sample images of pulled-to-sit examination. Infant in the video frames was being tested for neurological development at gestational age of 6 months at followup clinic of the Neonatology Department of Institute of Post Graduate Medical Education and Research (IPGME & R) and Seth Sukhlal Karnani Memorial (SSKM) Hospital, Kolkata, India [3].

with a client/server architecture. A simulation-based infant behavior analysis using a virtual environment is proposed in [14]. The authors have shown that unintentional physical injury to the children can be reduced using the proposed model. Video enabled tools are also used in the followup process for monitoring patients and their behavior [5], [12]. However, other than the visual recording system proposed by the authors of [6], no work has so far been reported in the literature that tries to automate the analysis of HINE. The above mentioned factors have motivated us to design video-based algorithms that can help the physicians conducting these examinations. Such algorithms will facilitate estimation of parameters that are used for analyzing of pulled-to-sit, lateral tilting, ventral suspension etc.

B. Video Object Tracking Algorithms and Their Limitations

In order to analyze Hammersmith examination recordings, it is often necessary to track individual body parts of the infant under examination. The problem of object tracking in videos has been extensively investigated. Many such tracking algorithms depend on block matching criteria like full search block matching (FSBM) [24] and variations of FSBM [22] such as diamond search (DS), directional diamond search (DDS), three step search (TSS), and new three step search (NTSS). These methods try to optimize the size of the search window based on various criteria. Alternatively, feature point-based tracking algorithms, where a feature point is selected and tracked by analyzing the motion of the block containing the feature point, have also been proposed [15]. Some of the problems encountered by the existing object tracking algorithms in videos are discussed as follows.

1) Feature point-based unsupervised algorithms [8], [9], [21] largely rely on detecting global features, which may not

[3]Goniometer is used in examinations like adductors and popliteal angle measurement.

be practical for all applications. For example, even a small error in locating feature points can degrade the performance of tracking. In addition, unsupervised algorithms are computationally time consuming and often do not meet timing constraints for some real life applications.

2) A feature point in one frame may match with multiple points in the subsequent frame, resulting in ambiguity in feature correspondence. Though various algorithms have been proposed to resolve such ambiguity, these algorithms often use exhaustive search such as FSBM and correlation over a large neighborhood for correspondence. Such techniques also incur high computational overhead and may not be suitable for real time tracking. Even algorithms using optimized searching methods sometimes fail to track some feature points [19], [23].
3) Tested with the available HINE recording, it has been observed that block matching-based algorithms often fail to track points of interest. Since the infants are kept unclothed during the examinations, foreground objects tend to have similar color and texture. This leads to multiple matches and nonunique identification of the feature points in successive frames. A small error introduced at each step would propagate and cause the failure of the tracking process.
4) Feature point-based object tracking algorithms try to detect landmark points in every frame and relate these points based on some criteria assuming that the object is rigid. Though such techniques have been found to be successful for some application domains such as tracking soccer ball [15] and human blob [19], [23], these algorithms often incur additional computational overload due to identification of feature points in every frame.

In view of the above-mentioned drawbacks of the existing tracking algorithms, in this paper we propose a new feature point-based tracking algorithm that dynamically selects k best nodes at every step. The pruning algorithm reduces the computational time of the tracking algorithm. The process begins with marking the feature points in the first frame of a video and then tracking these points to obtain the best possible trajectory. The proposed algorithm has been used to track movements of head and torso of the infant in the HINE recording. Next, a geometrical model of the scene is used to analyze the pulled-to-sit event.

The rest of the paper is organized as follows. In Section II, camera and experimental setup, geometrical model of the pulled-to-sit examination, proposed feature point tracking algorithm, and pulled-to-sit event recognition are discussed. Results of the analysis with normal as well as marker-based videos are presented in Section III. Conclusion and future plans are discussed in Section IV.

II. Materials and Methods

This section presents a detailed description of the camera and environmental setup, geometrical model of the pulled-to-sit examination, proposed feature point tracking algorithm, and event recognition process.

A. Camera and Experimental Setup

Visuals recorded from a single view may not be sufficient for automatic analysis since HINE assessment involves measurement of angles between various body parts, detection of postures, understanding reaction of the baby, and tracking body parts. Thus, pulled-to-sit, trunk and head position in sitting, popliteal angle measurement are viewed in sagittal plane (side view) since these examinations can be analyzed from the perspective projection of the corresponding frame in XY plane assuming that the Z-axis of the real world coordinate is parallel to the optical axis of the camera. On the other hand, adductors angle measurement and arm protection can better be analyzed if recordings are taken from the top. There are some examinations like lateral tilting, ventral and vertical suspension, which can be analyzed if videos are recorded in coronal view.

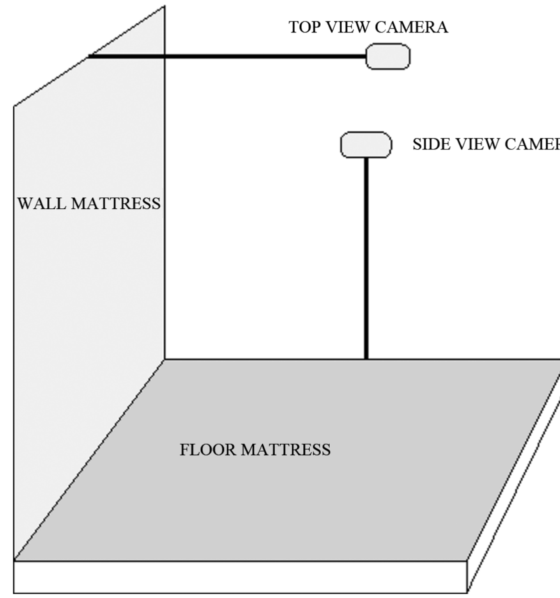

Fig. 2. Camera setup being used for recording videos and images of neurological examination process at followup clinic of SSKM hospital.

In view of this, the following camera and environmental setup has been used to record images and videos during HINE. A typical arrangement as shown in Fig. 2 has been used for data acquisition. Side view camera was used for recording sagittal as well as coronal view. We have used Sony XR500E video cameras. The cameras were focused in a close shot mode to minimize external interference. The cameras were interfaced with a personal computer through a USB TV Tuner (TECH-COM). A desktop computer powered by Intel Dual Core processor @2.13 GHz, 3 GB of DDR2 memory and Linux 2.6 kernel with gcc version 4.1.2 was used. We have coded the video capturing software using Intel openCV library. The software interface records videos of the examinations at 25 frames/s (fps) in Common Intermediate Format (CIF, dimension 352×288). To reduce the error during various image processing steps, examinations were conducted in a controlled environment with fixed illumination. The examination desk was covered with a soft mattress. This was done in accordance with the well established setup of the neuro-development clinic of the SSKM Hospital. Normally, all clothing is removed from the baby before HINE commences.

B. Geometrical Model of Pulled-to-Sit Examination

The side view camera records the scene of pulled-to-sit examination in sagittal view. In such a scenario, movement patterns

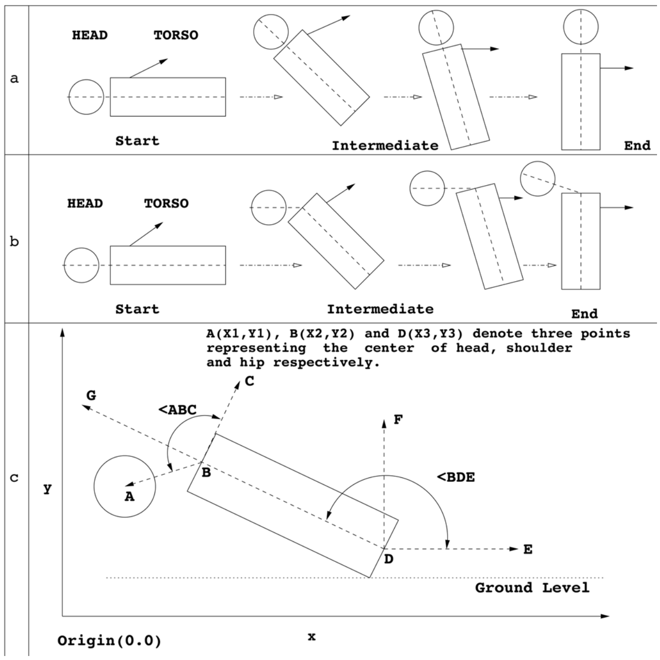

Fig. 3. Head-torso configuration for pulled-to-sit examination. (a) Model for normal head pulling. (b) Model for abnormal head pulling. (c) Proposed geometrical model for motion analysis.

of the baby can be modeled using the stick diagrams shown in Fig. 3(a) and (b). Relative movement of head and torso can be estimated in the 2-D projected space. Movement of the infant along the Z-axis (optical axis of the camera) can be ignored for this examination and the projection of the video frame on XY plane can be used for the analysis. It has been observed that during this examination, the bottom part of the torso (hip) works like a hinge and the entire body goes through a 90° rotation. In order to estimate relative movement between head and torso, one more hinge is introduced at the shoulder. When the baby lifts the head along the central axis of the torso, it is expected that the head will follow the torso and both will have a similar kind of motion as depicted in Fig. 3(a).

Suppose the center of head (A), shoulder (B), and hip (D), as shown in Fig. 3(c), are known at the beginning. Let the points A, B, and D be denoted by the coordinates $[X1, Y1]^T$, $[X2, Y2]^T$, and $[X3, Y3]^T$ with respect to the image coordinate system that is centered at (0,0). Using the reference coordinate system with origin as mentioned in the figure, a motion analysis of the torso section can be done as follows. First, a translation quantified by $[-X3, -Y3]^T$ is applied to the object. This shifts the point D to origin. Since it is assumed that during pulled-to-sit examination, hip (D) works as a hinge, the above-mentioned translation simplifies the geometry of the model. The translation equations are given in (1) and (2) where $T_x = -X_3$ and $T_y = -Y_3$ are translations along X and Y axis in the camera coordinate system

$$x' = x + T_x \tag{1}$$

$$x' = y + T_y. \tag{2}$$

Using the translation matrix in (3), (1) and (2) can be rewritten as (4). Thus, coordinates of A, B and D after such transformation can be computed using the translation described in

$$T_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -X_3 & -Y_3 & 1 \end{bmatrix} \tag{3}$$

$$[x', y', 1] = [x, y, 1] \times T_1. \tag{4}$$

Once the coordinates of the feature points are updated, motion of the torso can be analyzed. Assuming that the line segment denoted by $\overline{BD}$ represents the central axis passing through the torso of the baby, its angular velocity with respect to D can be estimated. Suppose $\overline{BD}$ makes an angle $\theta(\angle BDE)$ with the positive X-axis. At time $t = t_i$, gradient of $\overline{BD}$ is $\tan\theta_i$ and its rate of change can be estimated using

$$\Delta m_1^{i+1} = \frac{|\delta(\tan\theta_{i+1} - \tan\theta_i)|}{\delta(t_{i+1} - t_i)}. \tag{5}$$

This information is used to describe torso movement. It is assumed that during pulled-to-sit, the head-torso section $(\overline{AB})$ goes through an independent rotational movement around point D and the force applied directly to the torso $(\overline{BD})$ does not get transferred to the head-torso section. In such a scenario, relative movement of $\overline{AB}$ with respect to $\overline{BD}$ helps in analyzing the head movement. The following geometrical transformations are applied to A (representing center of head) and B (representing shoulder or joining point between head and torso) to estimate the motion of $\overline{AB}$ with respect to $(\overline{BD})$. First, a translation amounting to $[-X2, -Y2]^T$ is applied to the object using (6) with the translation matrix given in (7)

$$[x', y', 1] = [x, y, 1] \times T_2 \tag{6}$$

$$T_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -X_2 & -Y_2 & 1 \end{bmatrix}. \tag{7}$$

Next, to align $\overline{BC}$ and $\overline{BG}$ along x-axis and y-axis, a rotation as in (8) is applied with the rotational matrix (R) given in (9). This simplifies the geometry required for the motion analysis of the head of the baby

$$[x'', y'', 1] = [x', y', 1] \times R \tag{8}$$

$$R = \begin{bmatrix} \cos(\theta - 90^\circ) & -\sin(\theta - 90^\circ) & 0 \\ \sin(\theta - 90^\circ) & \cos(\theta - 90^\circ) & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{9}$$

Head movement can be quantified using the above mentioned analysis. The rate of change of gradient of $\overline{AB}$ with respect to B is estimated using (10) assuming that $\overline{AB}$ makes $\alpha(\angle ABC)$ angle with the positive x-axis of the coordinate system centered at B

$$\Delta m_2^{i+1} = \frac{|\delta(\tan\alpha_{i+1} - \tan\alpha_i)|}{\delta(t_{i+1} - t_i)}. \tag{10}$$

Equations (5) and (10) are used for pulled-to-sit event recognition that is discussed in Section II-D.

C. Proposed Feature Point Tracking Algorithm

In this section, we present a dynamic programming and node pruning-based feature point tracking algorithm for analyzing

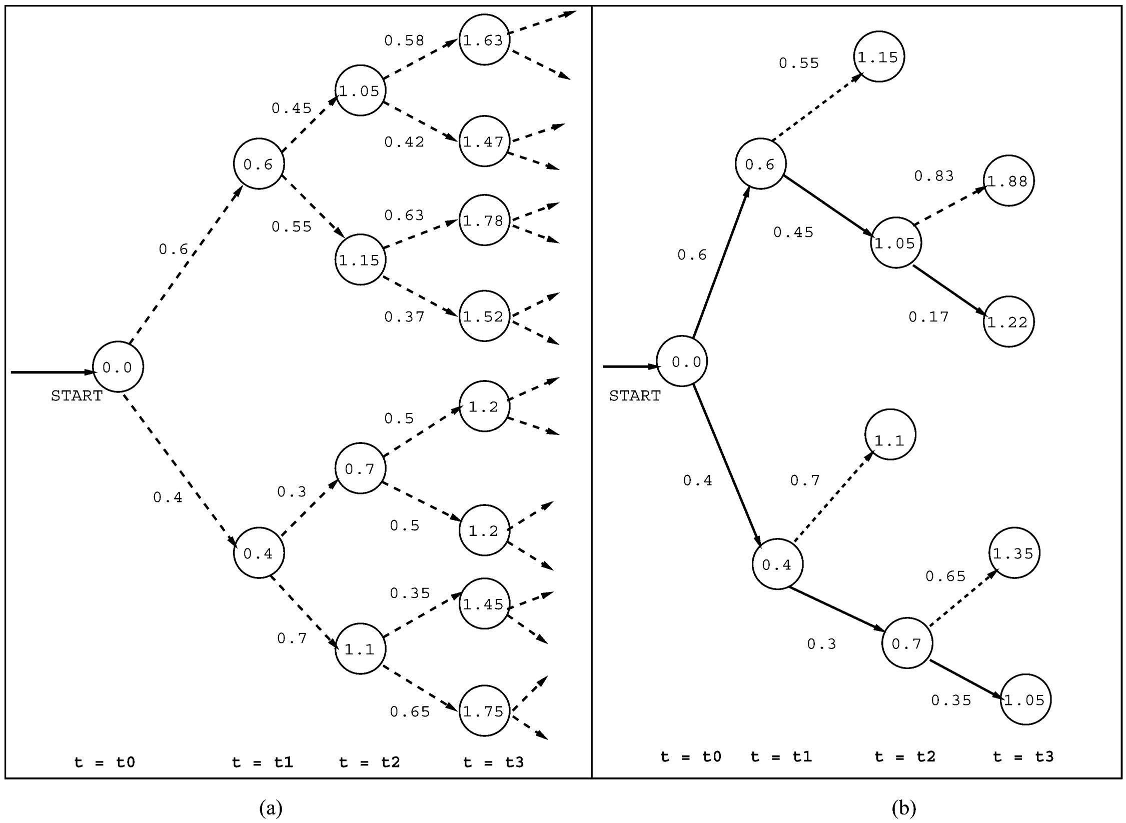

Fig. 4. Example of graph-based feature point tracking using dynamic programming and node pruning. (a) Exploration of all possible paths. (b) Exploration using node pruning with $k = 2$.

pulled-to-sit examination. The geometrical model described in the preceding section can be used only when the locations of head, shoulder, and hip are known. To accomplish this, a semi-automatic approach is adopted. At the beginning of the offline analysis, the examiner is asked to mark on the center of head, shoulder, and hip of the baby being examined. This is carried out with touch screen-based monitors with the physician marking these points using a finger tip. In addition, to increase the accuracy of the tracking algorithm, we have used white color markers (Micropore medical tape) to identify them. It has been observed that the best match using FSBM algorithm even fails to locate the exact position of the block containing the feature points in successive frames. This is because the foreground object (body of the unclothed baby) is homogeneous in terms of color and texture. It is possible that more than one surrounding blocks may have similar statistics. Thus, selecting the best match may not produce optimal result. Since the feature points are marked once, even a small error introduced in each step of prediction propagates throughout the entire tracking process. To overcome this problem, we propose a novel graph-based tracking algorithm that uses dynamic programming and node pruning to estimate possible locations of feature points in successive frames.

The proposed tracking algorithm has two passes. In the first pass, a graph is constructed where a landmark or feature point represents a node and inter frame connectivity between nodes represents edge. The graph is constructed as follows. The block containing the feature point in a frame is searched in the next frame within a surrounding window. The block that minimizes sum of absolute difference (SAD) value is assumed to be the probable location of the current block. Since such identification may introduce additive error, k nearest neighbor blocks are selected rather than selecting only the best one. Suppose a block x of the current frame matches with k nearest neighbor blocks of the next frame, say $y_1, y_2, \ldots, y_k$. The graph is constructed by connecting x with all the k blocks. The edges are assigned normalized weights $(d^x_{y_i})$ as mentioned in (11) and cost of reaching a node y from x is updated using (12)

$$d^x_{y_i} = \frac{SAD(x, y_i)}{\sum_{i=1}^{k} SAD(x, y_i)} \tag{11}$$

$$c(y_i) = c(x) + d^x_{y_i}. \tag{12}$$

When the algorithm explores all possible paths, it becomes highly computation intensive. An example of exploring all paths is shown in Fig. 4(a). Even with $k = 2$, the graph expands rapidly. Thus, for a complete pulled-to-sit examination, a total number of nodes in the graph becomes 2^{125} (average event duration 5 s and 25 fps video). Even if we process at a very low frame rate, say 5 fps, the number of nodes still remains in the order of 2^{25}. In view of this, we have adopted a pruning technique to reduce the total number of nodes that need be processed. The flow diagram of the algorithm is presented in Fig. 5.

The algorithm works as follows. Dynamically, we construct the graph such that at any level no more than k best nodes are added to it. At any given point of time, there are only k paths from the start node to the k nodes of the next frame. In each step, a maximum of k^2 additional nodes are checked. Without node

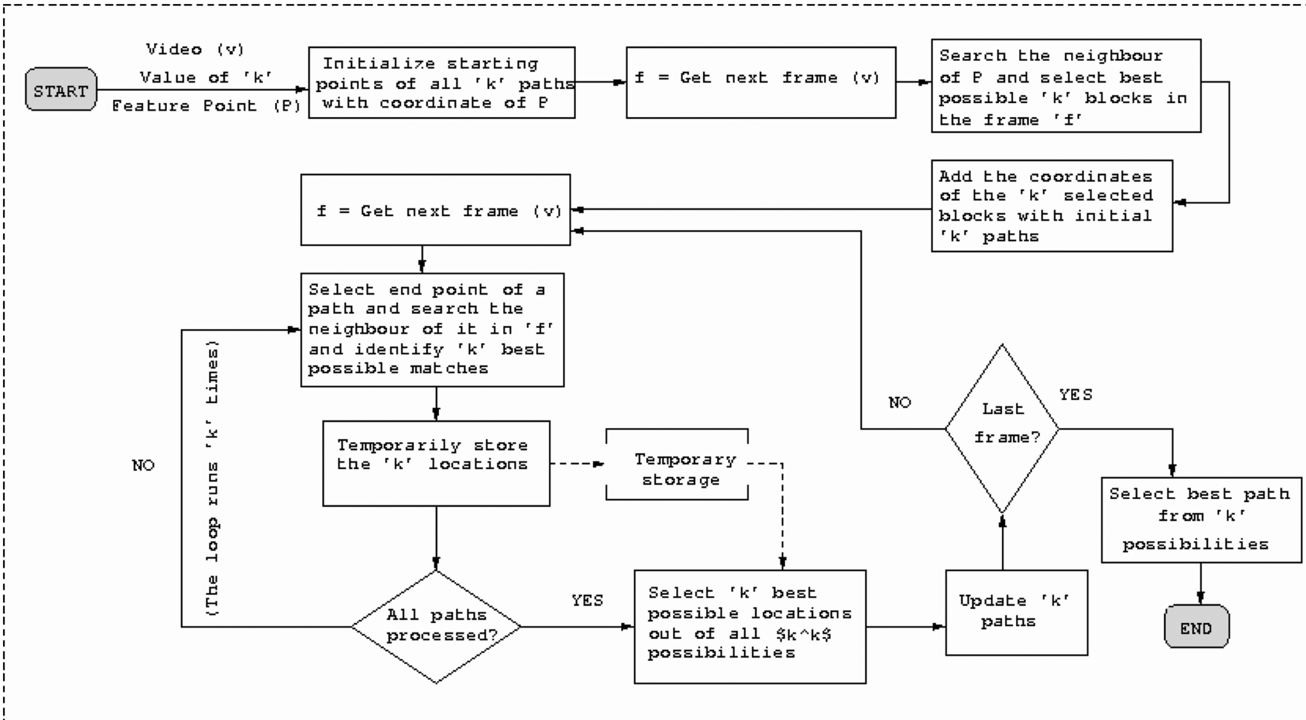

Fig. 5. Flow diagram of proposed feature point-based trajectory estimation algorithm. Algorithm computes minimum cost path out of k possible paths starting from initial location of a feature point being tracked.

pruning, the total number of nodes for processing p frames has been estimated to be $k^p - 1/k - 1$ which runs in k^p time. On the contrary, the total number of nodes that are processed using the proposed pruning algorithm becomes $[1 + k + (p-1)k^2]$ which runs in $O(pk^2)$ time. Since $k \ll p$ and normally its value is set to 2, the algorithm converges quickly. In Fig. 4(b), we have shown the paths that are explored with $k = 2$ using solid lines. Thus, for a video of 100 frames, the total number of nodes processed during the search is estimated to be 399. Finally, out of k paths, the smallest weighted path is selected as the probable route of the feature point being tracked.

D. Pulled-to-Sit Event Recognition

We now introduce a method to detect the events so that the HINE score related to pulled-to-sit examination can be estimated. In Table I, we present the conditions that are used to define events during pulled-to-sit examination. Equations (5) and (10) are used to characterize the events that may occur during this examination. Since the camera used to capture the videos of this examination has been fixed (height, zoom, and tilt of the side view camera are kept unchanged), it is possible to normalize the geometrical model of the scene described in Fig. 3. In addition, locations of the physician and the baby are kept unchanged. In such an environment, it can be assumed that the perspective view does not affect estimation of the relative angle based on the geometrical model of the scene. If the tracking algorithm follows the feature points accurately, a formulation of the event using (5) and (10) is possible. It has been observed that, at the beginning of the examination, i.e., at $t = t_0$, $\overline{AB}$, $\overline{BD}$, and $\overline{DE}$ line segments remain collinear to each other. Changes in angular velocity, e.g., Δm_1^{i+1} and Δm_2^{i+1} are recorded for successive frames. Finally, HINE score related to pulled-to-sit examination is estimated as follows.

- **Score 0:** Babies with abnormal development do not pull the head or even try to pull it (angle remains $> 30^\circ$ always). In such a scenario, a score of zero is awarded.
- **Score 1:** A score of one is given when more than once the baby pulls the head (according to the medical experts, in between successive frames the head might wobble within $\pm 30^\circ$).
- **Score 2:** According to the HINE chart [7], a score of two is not used for this examination.
- **Score 3:** If the condition as given in (13) is satisfied throughout the entire period of examination, i.e., the baby keeps the head aligned with the central axis of torso,[4] a score of three is given and the reaction of the baby for this examination is assumed to be normal. Value of T is estimated using the suggested permissible variation of relative angle between head and torso ($\pm 15^\circ$).

$$|\Delta m_1^{i+1} - \Delta m_2^{i+1}| < T. \tag{13}$$

When $-\tan^{-1} 15^\circ \leq T \leq +\tan^{-1} 15^\circ$, (13) can be rewritten using

$$\Delta m_1^{i+1} \approx \Delta m_2^{i+1} \tag{14}$$

III. Results and Discussion

The proposed approach for analyzing the recordings of pulled-to-sit examination has been tested with 43 babies (who

[4]According to the HINE experts, a variation of $\pm 15^\circ$ is acceptable and we have observed that if the locations of the camera and the baby undergoing examination are kept fixed, the perspective problem can be avoided

TABLE I
SCORES FOR GRADING INFANTS USING PULLED-TO-SIT EXAMINATIONS AND CORRESPONDING EVENTS. THRESHOLD VALUES HAVE BEEN DECIDED AFTER CONSULTING HINE EXPERTS OF THE HOSPITAL

Score	3	2	1	0
Description of Events	Head of the baby follows the torso ($\pm 15^o$ is acceptable)	Not used	More than once the baby pulls the head but the head may wobble within $\pm 30^o$	The head always remains below 30^o with respect to torso

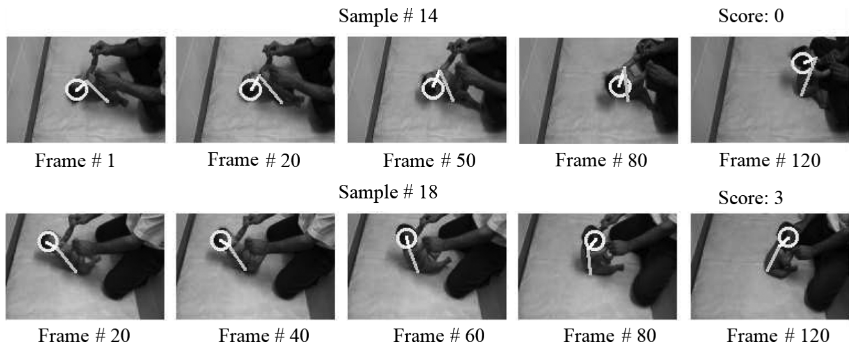

Fig. 6. Examples of successful detection of pulled-to-sit examination applied with two samples. Top row shows the event when baby (Sample #14) was unsuccessful in pulling its head during examination whereas the bottom row depicts a set of frames representing reaction of a normal baby (Sample # 18).

TABLE II
DESCRIPTION OF DATASET THAT HAS BEEN USED TO EVALUATE PERFORMANCE OF PROPOSED ALGORITHM

Gestational age at birth (in weeks) \ Corrected age during examination (in months)	3	6	9	12
Below 30 weeks	1	1	0	2
31 - 34 weeks	6	5	5	1
34 - 38 weeks	7	7	2	1
Above 39 weeks	2	0	2	1

went through this examination at the neuro-development clinic of SSKM Hospital during March 2010 to December 2010). In Table II, the summary of the infants covered during the experiments is presented. Since our primary focus was to automate the analysis of various examinations, the system has not been tested for longitudinal study. The samples are divided into four groups according to the corrected age. Sample size in each category is given in the table. Out of 43 cases, we have recorded videos of 13 babies with white colored markers. These videos are used to demonstrate how the results can be improved using markers.

First, we present the results obtained using the videos of 30 infants for which the videos were recorded without any markers. The examiner was asked to mark three locations, namely center of head, shoulder, and hip in the first frame of the video. Next, the tracking algorithm is applied to track the feature points in successive frames. The type of the event (as per Table I) during the pulled-to-sit examination is recognized using (13). The algorithm assigns a score based on the event type as mentioned in the preceding section. We have compared the scores with the physician's measurements. Table III presents the confusion matrix of semi-automatic measurement and physician's observation. Correctness of such classification results is often measured using statistical parameters like sensitivity and specificity. Sensitivity measures the proportion of true positives that are correctly identified. It is defined using (15) whereas specificity measures the proportion of negatives that are correctly identified (16). True positive (TP) is the count when a positive sample is labeled as positive, false negative (FN) is the count when a positive sample is labeled as negative, true negative (TN) is the count when a negative sample is labeled as negative and false positive (FP) is the count when a negative sample is labeled as positive. The values of TP, FN, TN, and FP are computed with respect to the doctor's assessment in the context of HINE. Sensitivity as high as 80% and specificity as high as 89% have been observed in this case

$$\text{sensitivity} = \frac{TP}{TP + FN} \tag{15}$$

$$\text{specificity} = \frac{TN}{TN + FP} \tag{16}$$

Sample frame sequences of two events with corresponding pulled-to-sit scores estimated using the proposed method are presented in Fig. 6. We have drawn a circle around the head of the baby and joined the lines connecting head, shoulder, and hip for better visualization. In the figure, one sequence (Sample #14) represents the reaction of a baby of 3 months and 5 days of corrected age at the time of examination and the other sequence (Sample #18) denotes the reaction of another baby of 6 months of corrected age at the time of examination. According to the expert's assessment, the baby of Sample #14 has less than normal neurological development as compared to the baby of Sample #18. The proposed method is successful in classifying them correctly.

A. Results Using Marker-Based Videos

The pulled-to-sit event recognition described in Section II-D may not be correct if the feature point tracking algorithm fails to precisely locate the points in successive frames. In order to reduce the chance of such errors, we have used color markers. During the test, micropore medical tape are placed on center of head, shoulder, and hip by the physician without causing any inconvenience to the baby. It is possible that the markers may get occluded due to the rotational movement of individual body

TABLE III
COMPARATIVE RESULTS BETWEEN SEMI-AUTOMATIC PULLED-TO-SIT EXAMINATION AND EXPERT'S MEASUREMENT CONSISTING OF 30 BABIES FOR WHICH VIDEOS WERE RECORDED WITHOUT PUTTING ANY MARKERS ON THEM

Estimated Measurement \ Expert's Measurement	Score: 0	Score: 1	Score: 3
Score: 0	3	0	1
Score: 1	0	5	4
Score: 3	0	1	16

TABLE IV
COMPARATIVE RESULTS BETWEEN MARKER-BASED PULLED-TO-SIT EXAMINATION AND EXPERT'S MEASUREMENT CONSISTING OF 13 BABIES

Estimated Measurement \ Expert's Measurement	Score: 0	Score: 1	Score: 3
Score: 0	3	0	0
Score: 1	0	1	0
Score: 3	0	1	8

Fig. 7. Baby with three visible markers attached to its body.

parts. However, we have observed that the input from the side view camera reduces the chance of occlusion to a large extent. A sample image of the marker-based examination is shown in Fig. 7. Table IV presents the result with marker-based videos of pulled-to-sit examination carried out at SSKM Hospital.

It has been found that the marker-based approach is successful in classifying all cases (sensitivity $\approx$ 92% and specificity $\approx$ 96%) except one when it wrongly classifies a baby with score 3 while the expert's assessment suggests the score should be 1. The results thus are encouraging since the presence of markers minimizes the error in tracking and makes the algorithm robust.

B. Failed Cases

Though the proposed approach produces encouraging results, in some cases the algorithm fails. For example, if the feature points are not tracked correctly, the motion-based analysis fails to describe the event correctly. One such example is shown in Fig. 8. It shows a scenario when the baby undergoing this examination was constantly moving, thus occluding the feature points

Fig. 8. Example of a failed scenario, when algorithm was unsuccessful to estimate score due to fast movements of body parts.

at various time instances making the estimation process erroneous.

C. Discussion

The video object tracking-based semi-automatic method for analyzing Hammersmith pulled-to-sit examination introduced in this paper can be used to develop biomedical technologies based on low cost systems. A wide section of the population can be covered if such tools reach remote areas. Similar work in this field [4], [6], [16] and clinical importance of the examinations [7], [10], [13], [17], [18], [20] have motivated us to design such semi-automatic tools.

Pulled-to-sit analysis using the proposed method has important significance in terms of clinical usability. The geometrical model of pulled-to-sit examination depends on the value of the threshold T mentioned in Section II-D. It has been observed that the threshold depends on the perspective projection of real world scene in camera coordinate. Thus, it is necessary to use a fixed camera setup to support threshold estimation in the context of pulled-to-sit examination. Other movement related examinations like lateral tilting, ventral and vertical suspension, which are relatively difficult to assess by visual observation, can be aided with software tools with appropriate examination specific geometrical models. It has also been observed that the pruning threshold k used in the feature point-based object tracking algorithm is sensitive to the type of video being used in tracking. It is experimentally found that a value of $k = 2$ works reasonably well on Hammersmith video recordings. It is evident that the true positive rate (sensitivity) has increased significantly (from 80% to 92%) when marker-based videos are used.

IV. Conclusion and Future Direction

Video-based low cost tools are useful in automating various examinations of the HINE tests. We have proposed a video object tracking-based method to automate Hammersmith pulled-to-sit examination. The dynamic programming and node pruning-based feature point tracking algorithm proposed in this paper can be used successfully to analyze pulled-to-sit examination assuming a geometrical model of the scene. It has been observed that the marker-based approach can improve the performance of the assessment process. The proposed approach can be used by trained paramedics as a decision support system for carrying out HINE assessment.

There are a few limitations of the current application which need to be addressed in future. Firstly, an extensive field trial is necessary for validation and standardization of this tool. We are also working towards developing algorithms for dealing with lateral tilting and ventral suspension types of Hammersmith examinations with the help of 3-D tracking. In such cases, depth estimation with multiple camera views is required.

Acknowledgment

The authors would like to thank Dr. A. Biswas of the School of Medical Science and Technology, IIT Kharagpur, India, and A. K. Ghosh of SSKM Hospital, Kolkata, India, for their valuable suggestions during the process of understanding the HINE assessment and its limitations.

References

[1] Calculating corrected age of infants Jul. 2011 [Online]. Available: http://depts.washington.edu/growing/Assess/Grca.htm

[2] Goniometer Jul. 2011 [Online]. Available: http://en.wikipedia.org/wiki/Goniometer

[3] Jul. 2011 [Online]. Available: http://www.ipgmer.gov.in

[4] J. Bhatt and G. Bebis, "Automatic recognition of a baby gesture," in *Proc. 15th IEEE Int. Conf. Tools Artificial Intelligence*, 2003, pp. 610–615.

[5] G. Bilodeau and S. Ammouri, "Monitoring of medication intake using a camera system," *J. Med. Syst.*, vol. 35, no. 3, pp. 1–13, 2009.

[6] D. Dogra, K. Nandam, A. Majumdar, S. Sural, J. Mukherjee, B. Majumdar, and S. M. Singh, "A user friendly implementation for efficiently conducting Hammersmith infant neurological," in *Proc. 12th Int. Conf. E-Health Networking, Applicat. Services*, 2010, pp. 374–378.

[7] L. Dubowitz, V. Dubowitz, and E. Mercuri, "The neurological assessment of the preterm and full term infant," in *Clinics in Developmental Medicine*. London, U.K.: Heinemann, 1981, vol. 9.

[8] T. Gao, G. Li, S. Lian, and J. Zhang, "Tracking video objects with feature points based particle filtering," in *Multimedia Tools and Applications*. Amsterdam, The Netherlands: Springer, 2011, pp. 1–21.

[9] B. Han, W. Roberts, D. Wu, and J. Li, "Robust feature-based object tracking," in *Proc. SPIE Conf. Algorithms for Synthetic Aperture Radar Imagery*, 2007, vol. 6568.

[10] L. Haataja, E. Mercuri, R. Regev, F. Cowan, M. Rutherford, V. Dubowitz, and L. Dubowitz, "Optimality score for the neurologic examination of the infant at 12 and 18 months of age," *J. Pediatrics*, vol. 135, no. 2, pp. 153–161, 1999.

[11] R. Kliegman, B. Stanton, J. Geme, N. Schor, and R. Behrman, *Nelson Textbook of Pediatrics*. New York: Elsevier, 2004.

[12] R. Logeswaran, "Cost effective patient location monitoring system using webcams," *J. Med. Syst.*, vol. 33, no. 5, pp. 399–407, 2009.

[13] A. Majnemer and B. Rossenblatt, "Prediction of outcome at school age in neonatal intensive care unit graduates using neonatal neurologic tools," *J. Child Neurol.*, vol. 15, no. 10, pp. 645–651, 2000.

[14] Y. Nishida, Y. Motomura, K. Kitamura, and H. Mizoguchi, "Infant behavior simulation based on an environmental model and a developmental behavior model," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, 2004, vol. 2, pp. 1555–1560.

[15] V. Pallavi, J. Mukherjee, S. Sural, and A. Majumdar, "Ball detection from broadcast soccer videos using static and dynamic features," *J. Visual Commun. Image Representation*, vol. 19, pp. 426–436, 2008.

[16] S. Singh and H. Hsiao, "Infant telemonitoring system," in *Proc. 25th IEEE Annu. lnt. Conf. Eng. Medicine Biology Soc.*, 2003, pp. 1354–1357.

[17] D. Romeo, M. Cioni, M. Scoto, L. Mazzonea, F. Palermoc, and M. Romeo, "Neuromotor development in infants with cerebral palsy investigated by the Hammersmith infant neurological examination during the first year of age," *Eur. J. Paediatric Neurol.*, vol. 12, pp. 24–31, 2008.

[18] D. Romeo, A. Guzzetta, M. Scoto, M. Cioni, P. Patusi, D. Mazzone, and M. Romeo, "Early neurologic assessment in preterm-infants: Integration of traditional neurologic examination and observation of general movements," *Eur. J. Paediatric Neurol.*, vol. 12, no. 3, pp. 183–189, 2008.

[19] C. Tomasi and T. Kanade, "Factoring image sequences into shape and motion," in *Proc. IEEE Workshop Visual Motion*, 1991, pp. 21–28.

[20] 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, and 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. Pediatrics*, vol. 105, no. 6, pp. 1216–1226, 2000.

[21] H. Wang and L. Zhang, "Object tracking based on local feature points," in *Proc. 3rd Int. Congr. Image Signal Processing (CISP)*, 2010, vol. 1, pp. 349–352.

[22] Y. Wang, J. Ostermann, and Y. Zhang, "Video processing and communications," in *Signal Processing Series*. Englewood Cliffs, NJ: Prentice Hall, 2002.

[23] C. Williams and M. Titsias, "Fast unsupervised greedy learning of multiple objects and parts from video," in *Proc. Conf. Computer Vision Pattern Recognition*, 2004, pp. 79–87.

[24] L. Y.-C. Lin and T. Shen-Chuan, "Fast full-search block-matching algorithm for motion-compensated video compression," *IEEE Trans. Commun.*, vol. 45, no. 5, pp. 527–531, May 1997.

Debi P. Dogra (S'09) received the B.Tech degree in computer science from Haldia Institute of Technology, Haldia, India, in 2001, and the M.Tech degree in computer science from Indian Institute of Technology (IIT), Kanpur, India, in 2003.

From June 2003 until July 2006, he worked as a Lecturer in the Computer Science Department of Haldia Institute of Technology, Haldia. From October 2006 until May 2007, he worked in the multimedia research team of Electronics and Telecommunication Research Institute, Daejeon, South Korea as a Researcher. Since July 2007, he has been a Research Scholar in the Department of Computer Science and Engineering in IIT Kharagpur. He has published more than 15 papers in international journals and conferences. His research interests are in the areas of image segmentation and video object tracking especially in the healthcare domain.

Arun K. Majumdar (M'74–SM'84) received the M.Tech. and Ph.D. degrees in applied physics from the University of Calcutta, Calcutta, India, in 1968 and 1973, respectively, and the Ph.D. degree in electrical engineering from the University of Florida, Gainesville, in 1976.

He is currently the Deputy Director of the Indian Institute of Technology (IIT) Kharagpur, West Bengal, India, and a Professor in the Computer Science and Engineering Department of IIT Kharagpur. He has held several administrative positions at IIT Kharagpur, including Dean (Faculty and Planning), Head of the Computer Science and Engineering Department, Head of the Computer and Informatics Centre, and Head of the School of Medical Science and Technology. He has been a member of several technical committees advising different Ministries of the Central and State Governments, as well as industries and institutions

in India on information technology-related matters. Before joining the IIT Kharagpur in 1980, he served as a Faculty Member at the Indian Statistical Institute, Calcutta, and Jawaharlal Nehru University, New Delhi. He was a Visiting Professor in the Computing and Information Sciences Department of the University of Guelph, Guelph, ON, Canada, during 1986–1987 and at George Mason University, Fairfax, VA, in 1999 and 2003. He has published more than 170 research publications in international journals and conferences and has supervised several Ph.D. and Masters' level theses and a large number of research and consultancy projects sponsored by different agencies and industries. His research interests include data and knowledge-based systems, multimedia systems, medical information systems, and design automation.

Dr. Majumdar is a Fellow of the Indian National Academy of Engineering and a Fellow of the Institution of Engineers (India).

Shamik Sural (M'93–SM'06) received the B.E. degree in electronics and telecommunication engineering from Jadavpur University, Calcutta, India, in 1990, the M.E. degree in electrical communication engineering from the Indian Institute of Science, Bangalore, India, in 1992, and the Ph.D. degree from Jadavpur University in 2000.

He is a Professor at the School of Information Technology, Indian Institute of Technology (IIT), Kharagpur, West Bengal, India. Before joining IIT, he held technical and managerial positions in a number of organizations both in India as well as in the U.S. He has published more than 120 research papers in reputed international journals and conferences. His research interests include database security, data mining, and multimedia database systems.

Dr. Sural has served on the Program Committee and Executive Committee of a number of international conferences, including the International Conference on Image Analysis and Recognition, the International Database Engineering and Applications Symposium, the IEEE Conference on Fuzzy Systems, and the International Workshop on Distributed Computing, and others. He served as the Chairman of the IEEE Kharagpur Section in the year 2006.

Jayanta Mukherjee (M'90–SM'04) received the B.Tech., M.Tech., and Ph.D. degrees in electronics and electrical communication engineering from the Indian Institute of Technology (IIT), Kharagpur, in 1985, 1987, and 1990, respectively.

He joined the faculty of the Department of Electronics and Electrical Communication Engineering at IIT, Kharagpur, in 1990, and later transferred to the Department of Computer Science and Engineering where he is presently a Professor. He served as the head of the Computer and Informatics Center at IIT, Kharagpur, from September 2004 to July 2007 and is presently serving as the Head of the Department of Computer Science and Engineering and the School of Information and Technology. He was a Humboldt Research Fellow at the Technical University of Munich in Germany for one year in 2002. He also has held short-term visiting positions at the University of California, Santa Barbara, University of Southern California, and the National University of Singapore. His research interests are in image processing, pattern recognition, computer graphics, multimedia systems and medical informatics. He has published more than 150 papers in journals and conference proceedings in these areas.

Dr. Mukherjee received the Young Scientist Award from the Indian National Science Academy in 1992. He is a fellow of the Indian National Academy of Engineering (INAE).

Suchandra Mukherjee received the MBBS degree from R. G. Kar Medical College, Kolkata, India, in 1993, and the MD degree specialization in pediatrics from M. K. C. G. Medical College, Berhampur, India, in 2000.

Currently she is an Associate Professor and in charge of the Neuro-Developmental Follow up Clinic for High-Risk Neonates, Department of Neonatology, Institute of Post Graduate Medical Education & Research (IPGMER), S. S. K. M. Hospital and the Associated Medical College, Kolkata, India. Prior to joining IPGME & R, she worked as a Senior Resident Registrar, Pediatrics, at Kothari Medical Centre, Kolkata, India, and as a Senior Registrar in the Department of Neonatology at Assembly of God Hospital & Research Centre, Kolkata, India. Since August 2003, she is associated with IPGME & R. She has been trained in assessment of development of children, Bayley Scale and toddler development at UCL, London in the year of 2011.

Dr. Mukherjee is a life member of Indian Academy of Pediatrics and National Neonatology Forum.

Arun Singh received the MBBS degree from Nilratan Sircar Medical College, Kolkata, India, in 1983, and the DCH degree from Calcutta Medical College, Kolkata, India, in 1987. He then received the MD degree specialization in pediatrics from All India Institute of Medical Science (AIIMS), New Delhi, India, in 1999.

Currently, he is an Associate Professor and Head of the Department of Neonatology, Institute of Post Graduate Medical Education and Research (IPGMER), S. S. K. M. Hospital and the Associated Medical College, Kolkata, India. Prior to joining IPGME &R, he was a Resident (Junior and Senior) in the Department of Pediatrics AIIMS. He has undertaken various projects like Concept of Sick Newborn Care Units at District and subdistrict levels and Development of Sick Newborn Care Units at a number of District Hospitals and Subdivision Hospitals under the Government of West Bengal, Rajasthan, Bihar, and, Andaman and Nicobar Islands. He has been associated with a number of collaborative research projects with various government agencies and institutes like Society of Applied Studies, Indian Council for Medical Research, Department of Science and Technology, Government of West Bengal, National Institute for Cholera and Enteric Diseases, Kolkata and IIT Kharagpur. He has published over 150 research papers in national and international journals and conferences. He has co-authored two books.