Med Biol Eng Comput (2014) 52:759–772
DOI 10.1007/s11517-014-1178-2

ORIGINAL ARTICLE

Video analysis of Hammersmith lateral tilting examination using Kalman filter guided multi-path tracking

Debi Prosad Dogra · Vishal Badri · Arun Kumar Majumdar · Shamik Sural · Jayanta Mukherjee · Suchandra Mukherjee · Arun Singh

Received: 7 April 2013 / Accepted: 12 July 2014 / Published online: 6 August 2014

Abstract Video object tracking plays an important role in many computer vision-aided applications. This paper presents a novel multi-path analysis-based video object tracking algorithm. Trajectory of the moving object is refined using a Kalman filter-based prediction method. The proposed algorithm has been used successfully to analyze one of the complex infant neurological examinations often referred to as Hammersmith lateral tilting test. This is an important test of the infant neurological assessment process, and this test is difficult to grade by visual observation. It has been shown in this paper that the proposed video object tracking algorithm can be used to analyze the videos of fast moving objects by incorporating application-specific information. For example, the proposed tracking algorithm can be used to assess lateral tilting test of the Hammersmith infant neurological examinations. The algorithm has been tested with several video recordings of this test which were captured at the neurodevelopment clinic of the SSKM Hospital, Kolkata, India during the period of the study. It is found that the proposed algorithm is capable of estimating the score for the test with high values of sensitivity and specificity.

Keywords Infant neurological examinations · Lateral tilting · Video object tracking · Multi-path tracking

D. P. Dogra (✉)
School of Electrical Sciences, IIT Bhubaneswar, Bhubaneswar 751013, India
e-mail: dpdogra@iitbbs.ac.in

V. Badri
Electronics and Radar Development Establishment, DRDO, Bangalore, India

A. K. Majumdar · J. Mukherjee
Department of Computer Science and Engineering, IIT Kharagpur, Kharagpur 721302, India

S. Sural
School of Information Technology, IIT Kharagpur, Kharagpur 721302, India

S. Mukherjee · A. Singh
NICU, Institute of Post Graduate Medical Education and Research, SSKM Hospital, Kolkata 700020, India

1 Introduction

Image and video processing is an integral part of medical image analysis and computer-aided healthcare services. Analysis of images and videos is critically dependent on having efficient techniques for image segmentation, object recognition, and object tracking. Thus, for analysis of medical images such as magnetic resonance (MR) or ultrasound (USG) images, efficient algorithms have been developed for segmenting region of interest and abnormality detection [19, 26]. Such algorithms provide considerable help to the radiologists. Similarly, echocardiogram videos can be analyzed to track the behavior of heart valves or objects of interest such as ventricles and aorta as described by the authors of [23]. In addition to that, video and image analysis-based techniques find application in infant care systems. For example, Singh and Hsiao [24] have used safe, compact and noninvasive sensors to record the movements of a baby with a client/server-based approach. It is a telemonitoring system that can be used for remote surveillance of infants. Simulation of infant behavior based on environmental model proposed by Nishida et al. [18] is another example of a video-based technique that has found applications in healthcare services. Video object tracking has been used in other healthcare-related services [15, 30] too.

Video object tracking algorithms can be classified into two groups, namely unsupervised and supervised. Unsupervised algorithms usually search feature points based on statistical information [9, 10, 29]. For example, authors of [10] have reported that landmark points can be identified and tracked to follow the trajectory of a moving object. Similarly, Gao et al. [9] have proposed an object tracking algorithm that uses feature point-based particle filtering with improved Scale Invariant Feature Transform (SIFT) [16]. First, the key points are identified using SIFT and then tracked. On the contrary, feature points are either known or marked by the user in a supervised approach. Some of the popular supervised algorithms based on structural information such as corner [25], human blob [14, 20], and moving car [13] are popular in this context.

Video object tracking can be done using feature point correspondence between consecutive frames [28]. In such techniques, either a feature point is detected in the first frame and tracked in successive frames or association of probable inter-frame landmark points is determined. The key idea behind both of these approaches is that, given a target feature point in one frame, a corresponding point in the next frame that minimizes matching distance based on some criteria is searched. It has been reported in the literature [9, 11, 17] that the existing feature point-based tracking algorithms often fail to precisely detect the location of an object in successive frames due to error accumulation. Especially, if the surrounding blocks of the object of interest have similar color and/or texture, determining the exact location of a block becomes erroneous. In such cases, error accumulation gets reduced considerably when multiple surrounding blocks of similar features are selected instead of picking a single block. However, for a video sequence of reasonably long duration, selection of multiple surrounding blocks of a feature point will increase the computational time. Therefore, we have proposed an efficient Kalman filter guided multi-path object tracking algorithm to overcome this.

We have analyzed the movement of body parts of infants undergoing Hammersmith infant neurological examinations (HINE) [8] using the proposed video object tracking algorithm. HINE is a set of quantitative methods for assessing neurological development of infants between 2 and 24 months of age. The examination set includes assessment of cranial nerve functions, posture, movements, tone, reflexes/reactions, and behavior. Outcome of these tests is used for early detection of neurological and developmental disabilities in pre-term as well as term infants [21, 22, 27].

We have already addressed a few critical problems associated with the conventional HINE methodology in some of our earlier work. For example, we have shown that visuals of the examinations can be recorded and used for offline analysis [4, 7]. We have proposed computer vision-based methods for analyzing examinations such as adductors angle measurement and pulled-to-sit [5, 6] by adopting similar approaches. In the present paper, analysis of video recordings of one the most critical tests that is often referred to as lateral tilting is presented. It was observed that the present methodology of assessment of this test is highly subjective. For example, a physician may evaluate the reaction of an infant inconsistently. In addition to that, all the consulted experts may not agree to a particular score even for a single subject. Therefore, notable intra-observer (0.095) and inter-observer (0.089) variance were recorded in the manual evaluation process [3]. This can be avoided using the proposed computer vision-aided technique presented in this paper.

The rest of the paper is organized as follows. The next section describes the overall methodology in its different subsections. Results of tracking using video recordings of lateral tilting test on several subjects are discussed in Sect. 3. Finally, conclusion and future directions of the present work are discussed in Sect. 4.

2 Methods

This section describes the geometrical model and the proposed video object tracking algorithm in the context of HINE lateral tilting test. First, a 2D geometrical model suitable for analyzing lateral tilting test is presented. Next, the proposed Kalman filter-based multi-path tracking (MPT) algorithm is described. Finally, how these two can successfully be used in combination to automate the analysis of lateral tilting is presented.

2.1 Description of the lateral tilting test

Lateral tilting test falls under the group of reflexes/reactions of the neurological examinations [8]. The test is carried out as follows. An examiner holds a baby around the pelvis or upper chest in a vertical posture and suddenly tilts to one side (90° rotation). The same has to be repeated by tilting the baby to the opposite side. Sometimes, a physician may tilt the baby from one side to the opposite side making a 180° rotation. After tilting, the physician holds the baby for about 3 s and observes the positions and movements of the head and limbs. The procedure is illustrated with some of the visuals of the test in Fig. 1. Possible reactions of the baby are depicted in four template images shown in the figure. After rotation, the physician observes the reaction of the head and upper limbs (upper leg and hand). Figure 1 depicts the reactions of the left hand, left leg, and head during the left to right tilting. Visuals of this examination are recorded using a single camera from the coronal view. In such a scenario,

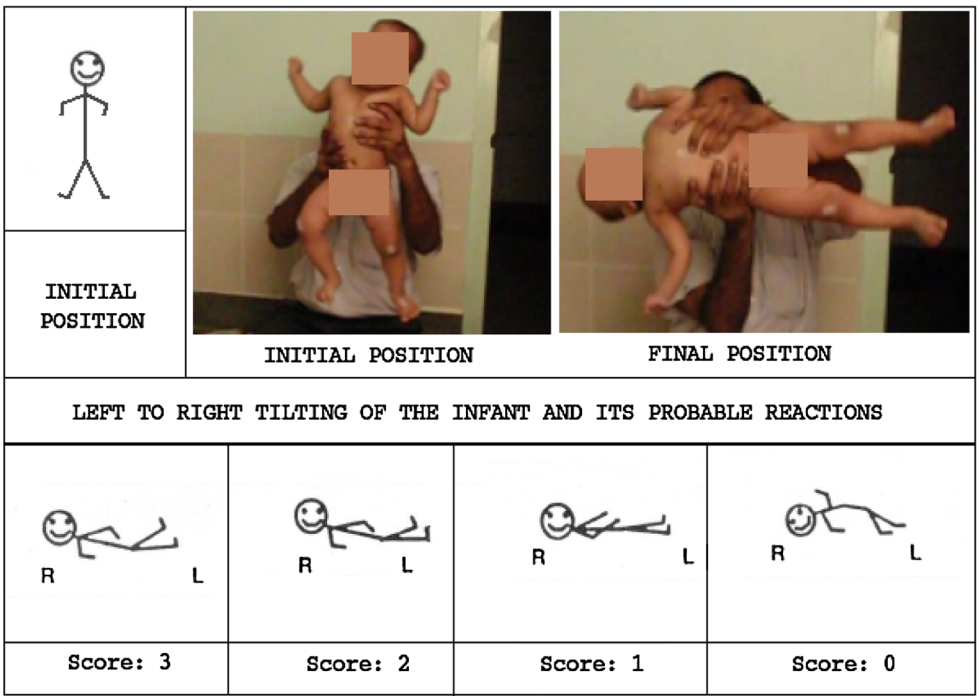

Fig. 1 Template images of the lateral tilting examination are shown in the *bottom row*. Two sample images are shown to provide better visualization to the reader

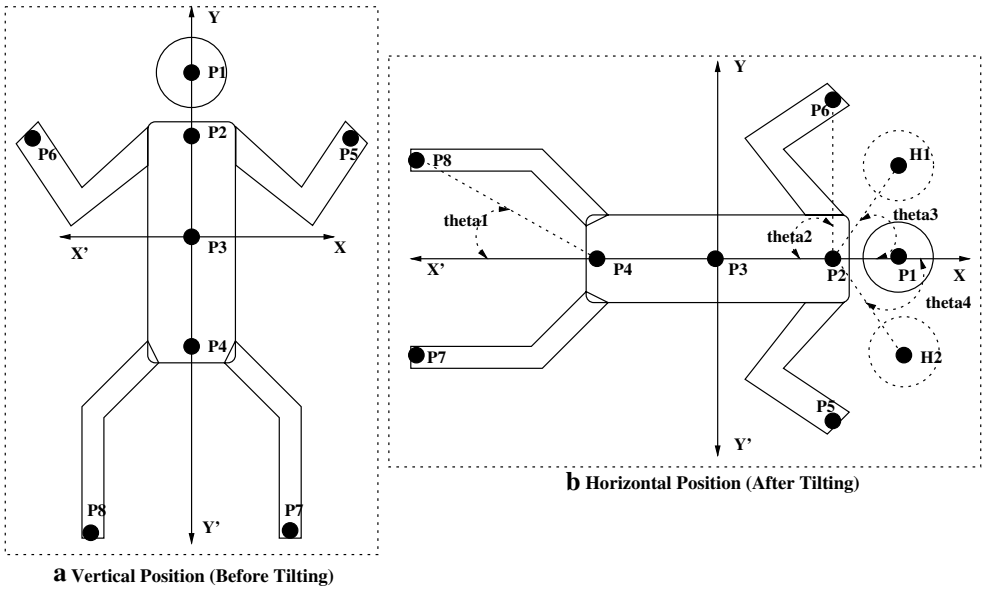

Fig. 2 Stick diagrams representing the 2D geometrical model of the lateral tilting examination in the *coronal view*, **a** posture of the baby before the tilting, **b** posture of the baby after tilting in the *left* direction (*anticlockwise*)

a 2D image of the scene is projected on the *XY* plane where *Z* axis is parallel to the optical axis of the camera. In the projected view, the rotation of the baby is assumed to be on the *XY* plane around *Z*. In such a configuration, the head pose of a normal baby is expected to be projected as shown in the template images. HINE scoring pattern corresponding to such orientations is shown in the figure.

2.2 Geometrical model of the lateral tilting examination

In order to track the head, center of the torso, upper leg, and upper hand, a 2D geometrical model as shown in Fig. 2 has been used. In the figure, two stick diagrams are depicted to demonstrate the position and orientation of the baby undergoing this test. It is assumed that the baby is rotated toward left (clockwise with respect to the optical axis of the

camera) of the physician. The model can easily be generalized to analyze the movements in the opposite direction.

Figure 2a shows the possible 2D projection of the posture prior to the rotation (vertical position) whereas Fig. 2b depicts one of the possible postures of the baby after it is rotated. Movements of the body parts in 3D world coordinate can be approximated using the features extracted from the 2D model. Let the points P_1 to P_8 represent various body parts of the infant that need to be tracked to analyze the test. Reflex movements of the head and limbs with respect to the center of rotation can be estimated based on the variation of the projected angle in 2D. For example, to detect the head pose as given in the chart shown in Fig. 1, the value of θ_3 (or $-\theta_4$) can be used. If the baby is normal, its head should always remain above the central axis ($\overline{P_2P_4}$) passing through the torso. In such cases, the angle θ_3 (in anticlockwise direction) can be useful. However, an abnormal baby may not lift its head even after the rotation is over, thus it will create $-\theta_4$ angle with the torso. Similarly, movements of the upper hand (left hand) and upper leg (left leg) can be estimated with respect to the shoulder joint (P_2) and hip joint (P_4), respectively. If the baby lifts the left leg, θ_1 is expected to increase and when there is a movement of the left hand, θ_2 increases. However, if P_8 or P_6 falls below the torso line ($\overline{P_2P_4}$), it can be assumed that the baby is unable to lift the upper limbs against the gravitational force which is a sign of abnormality. During the rotation in the reverse direction, reactions of head, right hand, and right leg are noted.

2.3 Assessment of lateral tilting

During the first half of the lateral tilting examination, the reflex motor action is not significant and the body as a whole goes through a rotational motion and acts like a rigid object. Reflex movements of the individual body parts of the baby that are guided by the neural response become dominant after sudden tilting is completed. The assumption helps in simplifying estimation of the reflex action which in turn enables the assignment of the scores as depicted in Fig. 1. The features mentioned in the preceding section can be used for assessing lateral tilting test on a scale of 0–3 as shown in the figure. In this quest, the events associated with the extent of movements of head, legs, and hands that are listed in Table 1 can be used for estimating the score based on angle measurement of various body parts. Since the reaction of the baby is observed for about 3 s after the tilting is completed, highest value of the angle during this period is recorded and used in the assessment.

Neurologists grade this test based on the reaction of head as well as limbs. However, higher weight is assigned to the head movement as compared to the limb movement. Five physicians of the neurodevelopment clinic of the SSKM Hospital, Kolkata were consulted while deciding the weights of head and limb movement in the final score estimation. Though, there exist no well-documented guidance for deciding these weights; however, all of the physicians consulted in SSKM Hospital have agreed that they usually give more weight to head movement than limb movement while assessing this test. Thus, based on their experience in this field, physicians of SSKM have decided to give 80 % weight to head movement and the remaining 20 % is evenly distributed among various limb movements. To capture the extent of tilting toward right or left during rotation in both sides, the weight assigned to head movement is equally divided into two parts. Thus, the final score (S_{final}) is estimated using (1), where w_1 and w_2 are computed from the percentage of weights of the individual components. Therefore, normalized weight of head movement (one direction) is taken as 0.4 and normalized weight for hand and leg movement is set to 0.05.

$$S_{\text{final}} = \Big[w_1 S_{\text{head}}^{\text{right tilt}} + w_1 S_{\text{head}}^{\text{left tilt}} + w_2 (S_{\text{LH}} + S_{\text{LL}} + S_{\text{RH}} + S_{\text{RL}}) \Big] \tag{1}$$

2.4 Proposed tracking algorithm

The events listed in Table 1 can be detected if the points representing the body parts shown in the 2D geometrical model are tracked successfully during the test. To track these points, a novel algorithm has been proposed in this section. It is referred to as the MPT algorithm that is guided by a Kalman filter-based prediction model. The algorithm is based on feature point association between successive frames. In the present context, Hue (H)–Saturation (S)–Intensity (V) color space has been found to be convenient.

Table 1 Scores based on independent movements of head, left hand, left leg, right hand, and right leg

Features	Scores			
	Score: 3	Score: 2	Score: 1	Score: 0
Head	$\theta_3 \geq 20°$	$10° \leq \theta_3 < 20°$	$0° \leq \theta_3 < 10°$	$-\theta_4 \leq \theta_3 < 0°$
Left hand	$\theta_2 \geq 60°$	$30° \leq \theta_2 < 60°$	$0° \leq \theta_2 < 30°$	$\theta_2 < 0°$
Left leg	$\theta_1 \geq 30°$	$15° \leq \theta_1 < 30°$	$0° \leq \theta_1 < 15°$	$\theta_1 < 0°$
Right hand	$\theta_2 \geq 60°$	$30° \leq \theta_2 < 60°$	$0° \leq \theta_2 < 30°$	$\theta_2 < 0°$
Right leg	$\theta_1 \geq 30°$	$15° \leq \theta_1 < 30°$	$0° \leq \theta_1 < 15°$	$\theta_1 < 0°$

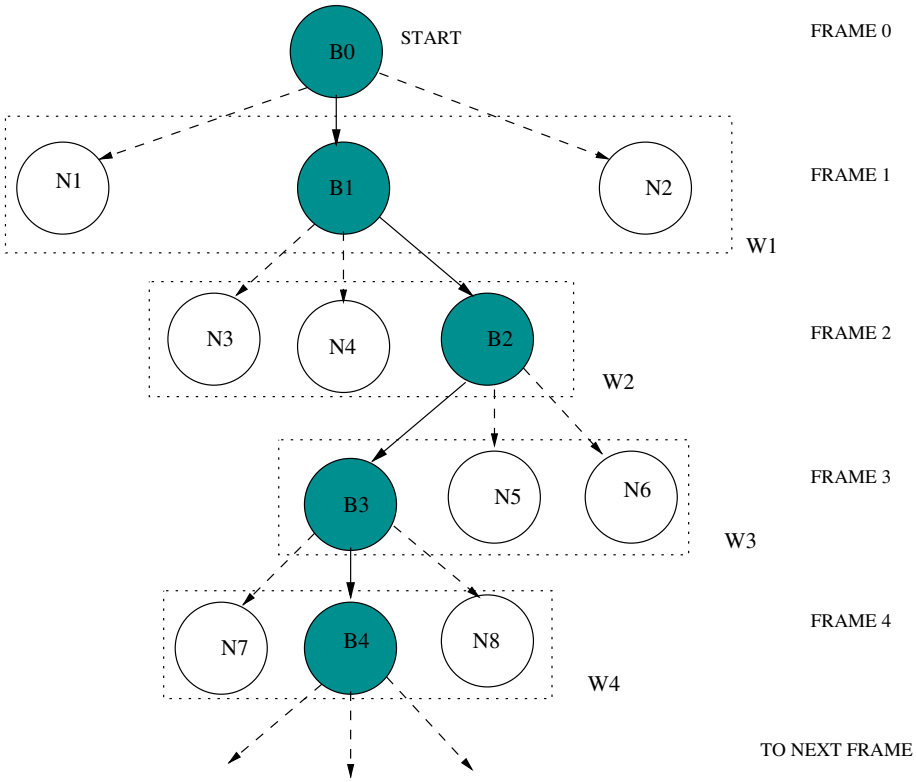

Fig. 3 An example of tracking with single path exploration based on RMS error

Root mean square (RMS) error is computed using (2), where $M \times N$ represents the block size, B_i and B_{i+1} represent the blocks in frames i and $i+1$, and $C = \{H, S, V\}$, respectively.

$$\mathrm{RMS}(B_i, B_{i+1}) = \frac{1}{M \times N} \sqrt{\sum_{m=1}^{M} \sum_{n=1}^{N} \sum [C_i(m,n) - C_{i+1}(m,n)]^2} \quad (2)$$

2.5 Single path exploration

The RMS error is estimated between the blocks of two successive frames. Based on this error measurement, trajectory of the block can be estimated. In single path mode, the block having the least RMS measure with the current block is selected. Probable path of a block (B_0) representing a body part being tracked using this method is shown in Fig. 3. The shaded blocks (B_0, B_1, B_2, B_3, B_4) represent the final trajectory of the point that has been assumed to be marked on the first frame of the video while the unshaded blocks (N_1, N_2, N_3, N_4 and so on) represent the neighboring nodes that are not included in the final trajectory. In each step, the location of the block is searched inside a window. For example, the window W_1 is used when the probable location of B_0 is searched in the next frame (FRAME 1). Next, RMS error between block B_0 and each of its surrounding nodes in the next frame is computed and the node having minimum RMS (B_1 in this case) is selected. The process advances to the next level and the same strategy is applied to locate B_2 and so on. This can be extended to multi-path analysis by selecting more than one probable surrounding blocks.

2.6 Refinement using Kalman filter

The tracking algorithm discussed in the preceding section uses minimum RMS-based prediction of the probable location of a block in successive frames. However, it has been observed that RMS-based prediction heavily relies on local feature. Such predictions may fail to accurately track an object if the feature used in the association is extracted from noisy measurements. In such cases, correction of the measurements can be done using Kalman filter. The prediction works in a typical two-phase alternate scheme based on *Predict* and *Update* discussed below [12]. Prediction is done for advancing the state until the next scheduled observation and the update is performed by incorporating the new observation. Suppose, $\boldsymbol{x}$ denotes the state variable of a n dimensional dynamical system described by the state transition equation at time $t = t_k$ such that $\boldsymbol{x}_k = F\boldsymbol{x}_{k-1} + w_k$ where F is state transition matrix and w_k is a zero mean Gaussian white noise process with autocorrelation matrix W. The measurement variable z_k is a linear combination of $\boldsymbol{x}_k$ and is given by $z_k = H\boldsymbol{x}_k + v_k$. Here, v_k is a zero mean Gaussian white noise process with autocorrelation R. $\hat{x}_{k|k}$ is the a posteriori state estimate at time k given observations up to and including at time k, and $P_{k|k}$ is the a posteriori error covariance matrix. Then according to

Kalman filter model [12], the predict and update steps are performed as follows where I is the identity matrix and $\mathbf{T}$ is the transpose operation:

Predict

Predicted (a priori) state estimate, $\hat{x}_{k|k-1} = F\hat{x}_{k-1|k-1}$
Predicted (a priori) estimate covariance, $P_{k|k-1} = FP_{k-1|k-1}F^{\mathbf{T}} + W$

Update

Innovation or measurement residual, $\tilde{z}_k = \boldsymbol{z}_k - H\hat{x}_{k|k-1}$
Innovation (or residual) covariance, $S_k = HP_{k|k-1}H^{\mathbf{T}} + R$
Optimal Kalman gain, $K_k = P_{k|k-1}H^{\mathbf{T}}S_k^{-1}$
Updated (a posteriori) state estimate, $\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k\tilde{z}_k$
Updated (a posteriori) estimate covariance, $P_{k|k} = (I - K_kH)P_{k|k-1}$.

Typically, the two phases alternate, with the prediction advancing the state until the next scheduled observation, and the update incorporating the observation. If an observation is unavailable at a particular time instance (the input data may not be available due to some error in the sample acquisition step), the update may be skipped for the current step and the prediction can be performed based on previous available data. The autocorrelation matrix of process noise and measurement noise are given in (3) and (4), where T is inter-frame duration. Value of w_0 was selected after several experiments (it was varied from 0.1 to 10) and r_0 was estimated using sample data.

$$W = \begin{bmatrix} T^3/3 & T^2/2 & 0 & 0 \\ T^2/2 & T & 0 & 0 \\ 0 & 0 & T^3/3 & T^2/2 \\ 0 & 0 & T^2/2 & 0 \end{bmatrix} w_0 \tag{3}$$

$$R = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} r_0 \tag{4}$$

Since the point is marked by the user, its initial position is known. Thus, corresponding elements of the state covariance matrix ($P_{0|0}$) are initialized with zero. However, the velocity of the object is not known at the beginning. An approach similar to [2] has been used and velocity estimates of the state vector are initialized to zero while corresponding elements of the state covariance matrix are set to a large value to reflect the uncertainty in the velocity estimation. In the present context, it has been found that an initial value of 10 produces satisfactory results. Thus, the matrix $P_{0|0}$ as in (5) is used in the recursion.

$$P_{0|0} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 10 \end{bmatrix} \tag{5}$$

One of the important parameters in Kalman filter-based prediction is the state transition matrix (F). In case of lateral tilting test, the state variables associated with a point (x, y) of any block can be taken as $\boldsymbol{x} = [x, \dot{x}, y, \dot{y}]^{\mathbf{T}}$ where $\dot{x}$ and $\dot{y}$ denote velocity. It is assumed that during the first phase of the tilting, an infant undergoes rotational motion around the body center with a constant turn rate Ω. In such a scenario, a state transition matrix as given in (6) can be used. Depending on the frame rate of the video, a good guess of the value of Ω can be made as discussed next. Suppose, average duration of the tilting is half a second and the baby is rotated 90° during this short interval. In such a scenario, if the frame rate is assumed to be f frames/s, then inter-frame gap becomes $1/f$ s. Thus, the amount of inter-frame rotation would be $180/f$ degrees approximately. In view of this, Ω may be considered as $180/f$ and the inter-frame duration (T) can be taken as $1/f$ s. For example, with $f = 25$, an estimate of $T = 40$ ms and $\Omega = 7°$ can be used. It was noted in [1] that in the presence of noise, Ω may vary with time and therefore can be estimated iteratively using $\Omega(t+1) = \Omega(t) + \eta_\omega(t)$. Such variation of Ω may introduce accumulation of errors in the estimation of state vectors if the Kalman prediction is repeatedly used from $x(t)$ to $x(t+k)$ for $k > 1$. However, if RMS-based prediction is carried out along with the estimated value of $x(t)$ after each step, such error accumulation can be avoided.

$$F = \begin{bmatrix} 1 & \frac{\sin(\Omega \times T)}{\Omega} & 0 & -\frac{1-\cos(\Omega \times T)}{\Omega} \\ 0 & \cos(\Omega \times T) & 0 & -\sin(\Omega \times T) \\ 0 & \frac{1-\cos(\Omega \times T)}{\Omega} & 1 & \frac{\sin(\Omega \times T)}{\Omega} \\ 0 & \sin(\Omega \times T) & 0 & \cos(\Omega \times T) \end{bmatrix} \tag{6}$$

Similarly, the state transition matrix as defined in (7) can be used for tracking objects having translational motion. For example, this can be used during the analysis of the reflex movements of the body parts after the tilting is over.

$$F = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{7}$$

RMS-based prediction is considered as the observation in single path tracking (without Kalman filter). Since at each step, location of the center of a block is updated based on the RMS, this is assumed to be the measurement at each step. Thus, if the Kalman filter-based correction is adopted, then each predict step can use the RMS guided result as observation. Next, in the update step, the measurement can be corrected. The updated value can be assumed as the corrected location of the block being tracked. This has been explained with an example shown in Fig. 4. The final path of the block B_0 is shown using solid lines. Kalman filter-based refinement of single path tracking algorithm is presented below.

Refinement algorithm

- Step 1: Let a point, say P, be marked on the first frame of a video.

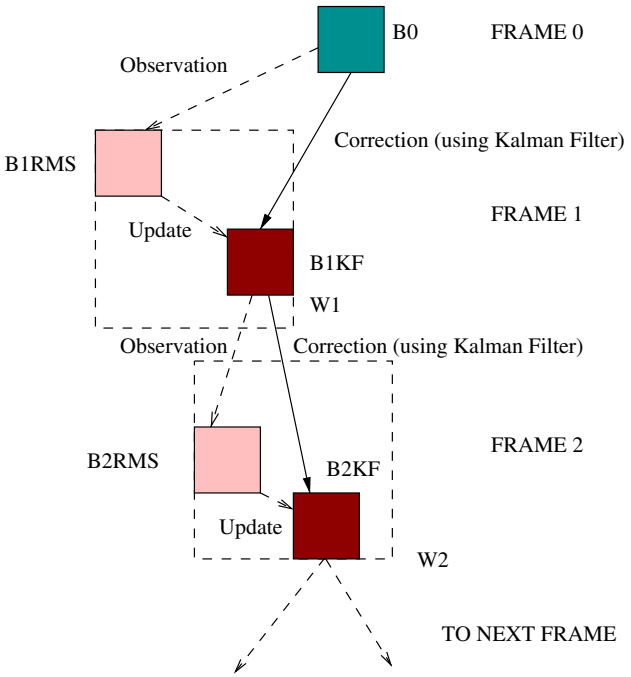

Fig. 4 An example of refinement of the block prediction using Kalman filter where B_i^{RMS} represents a block that has been predicted using minimum RMS criterion and B_i^{KF} denotes its corresponding updated location

- Step 2: *Prediction using RMS* Get the next frame of the video and search the window around P for a probable match according to the minimum RMS error. Let the predicted location of the point be represented by Q_t.
- Step 3: *Update using Kalman filter* Use Q_t as the observation at time t. The updated value $\hat{Q}_t$ is returned by the filter and this is assumed to be the final position of the block at time t.
- Step 4: Assign $P = \hat{Q}_t$ and go to step 2 to get the next observation Q_{t+1} using RMS-based prediction.

2.7 Limitations of single path tracking

The single path algorithm discussed in the preceding subsection has been used for tracking body parts of the infants during the lateral tilting test. Some of the sample frames extracted from a video recording are shown in Fig. 5. It is evident that the algorithm fails to track the leg of the baby from Frame 12 onwards. This is due to fast movement of the legs. Thus, the tracking algorithm discussed in the preceding section may not be appropriate in the present context.

The problems encountered are being illustrated with an example shown in Fig. 6. Let a best match for the block B_0 in the first frame is searched in the next frame. In a typical scenario, single path searching algorithm may select the block B_1 as the best choice according to the minimum RMS criterion and the block B_2 may not be selected in spite of a low RMS. It may be possible that in the next level, the

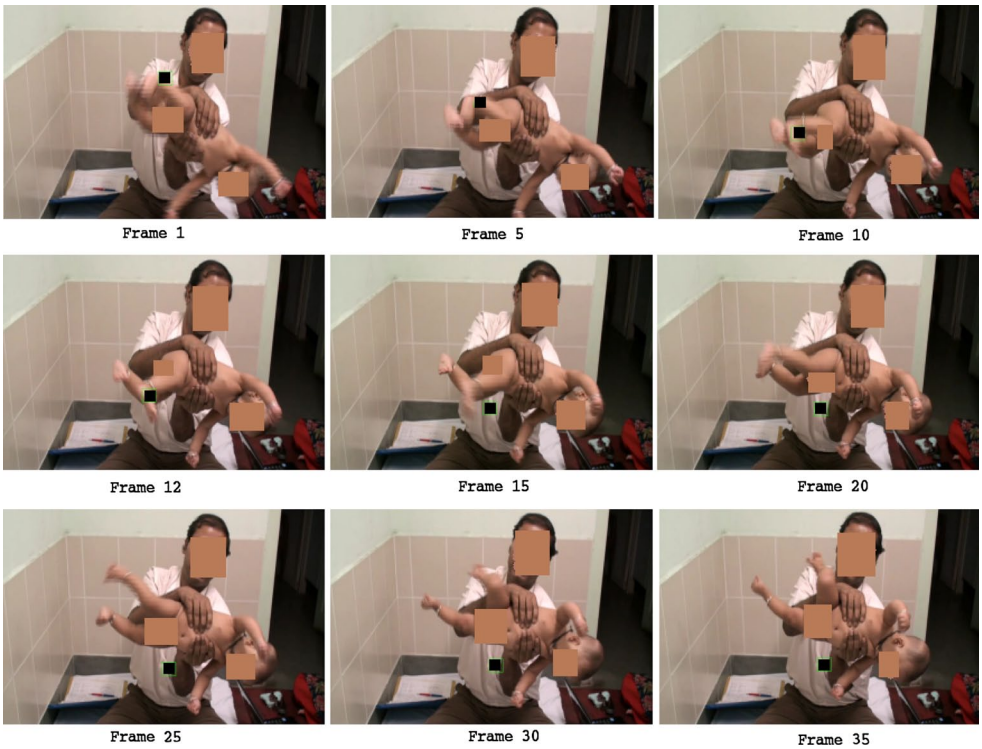

Fig. 5 Tracking of a body part using the single path-based approach

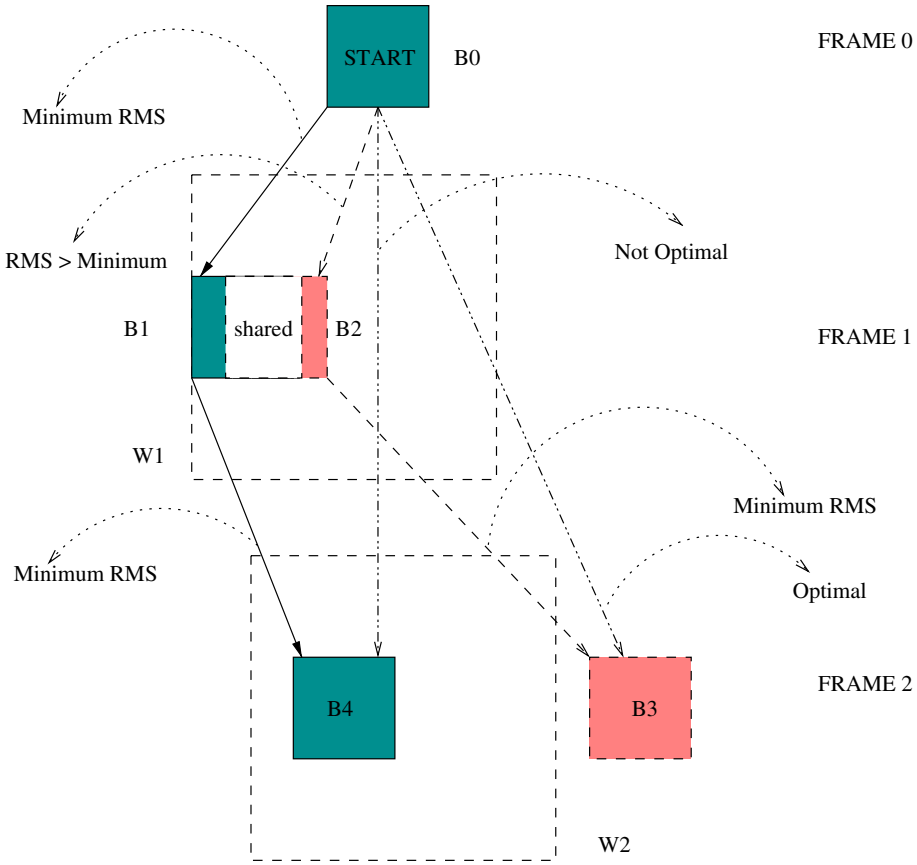

Fig. 6 Example of a situation when a multi-path strategy can result in better estimation of the trajectory. Final path is denoted using *solid lines* whereas unexplored paths are represented using *dashed lines*

RMS error between the blocks B_3 and B_0 is minimum. However, as B_2 is ignored in the previous step, the algorithm does not consider B_3 in the computation of the final path. On the other hand, in spite of not having the optimal RMS with B_0, B_4 gets selected since it is linked via B_1. In addition to that, the block B_3 may move outside the search window (W_2) due to fast movement of the object. In such situations, there is no scope to compute RMS between B_1 and B_3 at the second level. Thus, the best path algorithm (greedy approach) may not always guarantee an accurate trajectory.

2.8 Proposed multi-path tracking algorithm

Some of the limitations of the tracking algorithm based on single path exploration can be overcome using MPT approach. In MPT, instead of single path searching, multiple paths are explored and the probable trajectory of a body part can be estimated. Selection of multiple candidates is done based on the value of a threshold described later in this section. However, if MPT is carried out for a long period, it will be time consuming. To overcome this, a novel algorithm has been proposed that switches from the single path to MPT mode depending on the present observation or according to a given terminating criterion. The entire trace of the algorithm can be divided into multiple segments containing single path or MPT phase as shown in the Fig. 7. Switching between single path and MPT is done based on a threshold (T_{MPT}). The algorithm is described below.

MPT algorithm

- Step 1: ***Initialization:*** Let the point P denotes the center of the block representing a body part being tracked in the first frame of the video and $L_{\max}$ denotes the highest permissible level in MPT.
- Step 2: ***Single path testing:*** Get the next frame of the video and search the surrounding window for probable matching. If $\mathrm{RMS}(P,Q) < T_{\mathrm{MPT}}$ is true for any $Q \in \mathrm{neighbour}(P)$, select the particular Q having minimum RMS and repeat this step, else go to the next step.
- Step 3: ***MPT initialization:*** Start the MPT mode of the algorithm. Let $S = \{S_1, \ldots, S_K\}$ represent the set of K best possible nodes that are children of a node reached after Step 2 before entering MPT phase according to the least RMS criterion. Let L denotes the current depth of the tree inside MPT and be initialized to zero.
- Step 4: ***MPT recursion:*** Get the next frame of the video. Initialize the search around the K locations corresponding to blocks in S. For every element of the set S

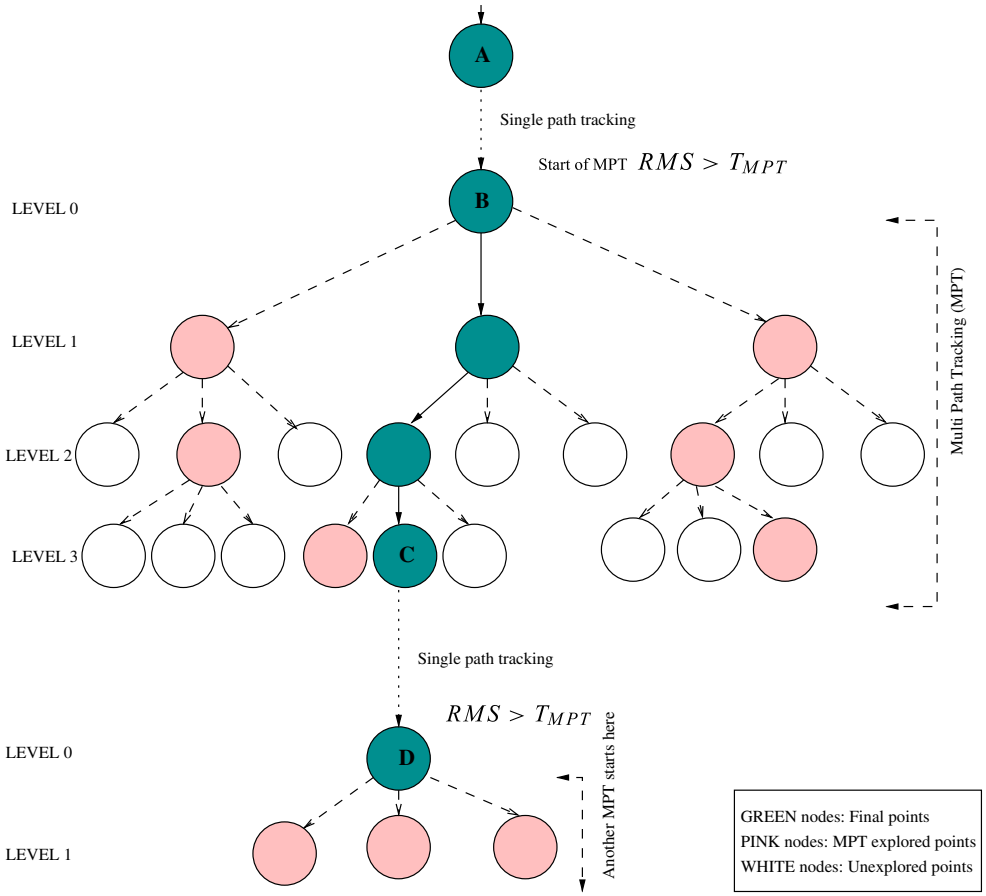

Fig. 7 An example of MPT algorithm. Value of the threshold T_{MPT} has been estimated after carrying several experiments on a number of sample videos of lateral tilting examination

, K best possibilities are selected from the list of its children. Thus, a set of K^2 possible locations are found. Let this set be represented by:

$$M_i^j = \{\{M_1^1, \ldots, M_1^K\}, \ldots, \{M_K^1, \ldots, M_K^K\}\}$$

- If $RMS(S_i, M_i^j) < T_{MPT}$ for any $1 \leq i \leq K$ and $1 \leq j \leq K$, select the particular block M_i^j that is having minimum RMS with its parent S_i and go to Step 2 (switch to single path tracking).
- Else,
  - If $L < L_{\max}$, select K best blocks (according to minimum RMS criteria) from the list M_i^j, make $L = L + 1$, and repeat Step 4 with $S_1, \ldots, S_K$ in S replaced by these blocks (continue MPT).
  - Else, select the best block M_i^j and the intermediate path from the starting point of the current MPT to M_i^j is added with the final path and go to Step 2 (MPT termination).

The algorithm enters into MPT mode when RMS error of all of the neighboring blocks is higher than the threshold T_{MPT}. In the MPT mode (step 3), instead of selecting the location having minimum RMS, K-nearest neighbor blocks are examined. Inside the MPT, each of the K blocks is examined separately. Thus, using the similar K-nearest neighbor approach, a total of K^2 blocks are examined in the next level and out of all such possibilities, only the best K options are kept. During this elimination phase, it is checked whether the RMS errors of all of the selected K blocks are above T_{MPT} or not. If the condition fails (i.e., at least one of the K blocks satisfies RMS $< T_{\text{MPT}}$), the algorithm switches to the single path mode. However, if the above mentioned condition is true in $L_{\max}$ successive frames, the algorithm automatically switches to the single path mode after selecting the least weight path of the MPT tree.

The algorithm is explained with an example shown in Fig. 7. Let the node A in the figure be the starting point of the tracking algorithm and it represents the block that has been marked or identified on the first frame of a video. Single path tracking is continued till node B and during this

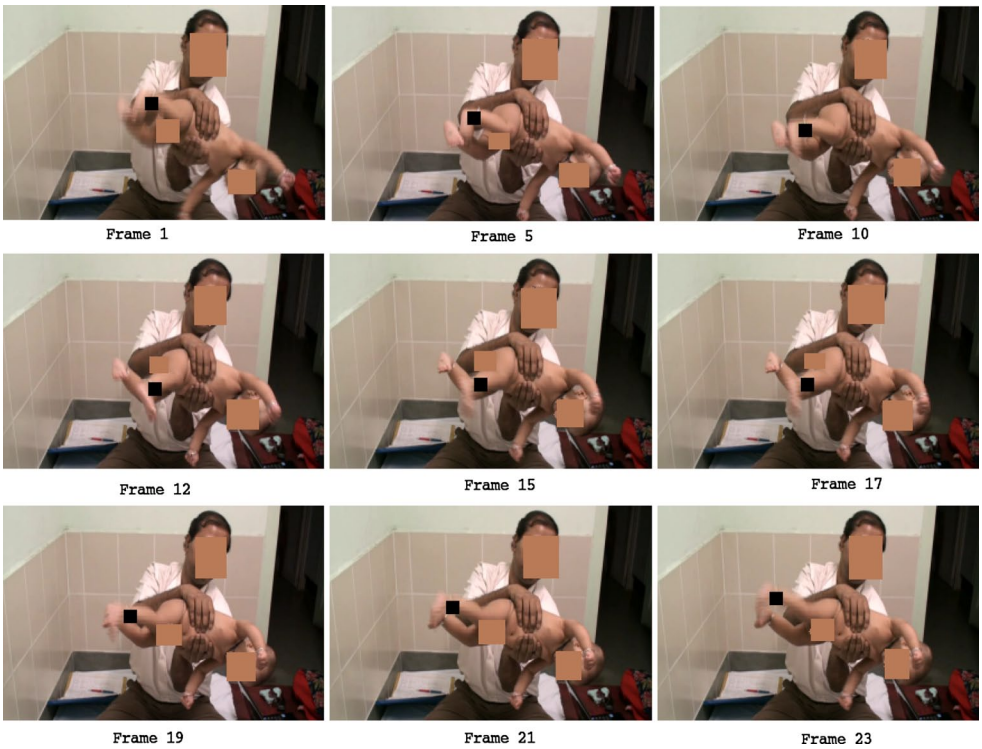

Fig. 8 Tracking of a body part using the proposed multi-path (MPT) algorithm

period, RMS error of at least one of the surrounding blocks remains below the threshold T_{MPT}. However, at B, the condition fails and the algorithm enters into the MPT mode and it is continued till node C. At each level, K^2 nodes are examined and only K best nodes are selected. However, since the MPT reaches the maximum allowable limit of the level (in this case, the highest level is set to 3), the best path up to node C is selected before it exits from the MPT mode. Nodes of the final path up to this point and intermediate nodes explored by the MPT algorithm are colored distinctly for clarity. White circles denote unexplored nodes. The algorithm runs in single path mode between the nodes C and D, and another MPT starts from D onwards. The RMS-based block prediction used in the MPT algorithm can be further refined using the Kalman filter guided correction scheme discussed in the preceding section. In the MPT mode, a threshold is applied to filter out points that have lower matching score with respect to the candidate node according to the RMS criteria. However, to minimize the computation load, no more than five surrounding nodes are selected in each step while the algorithm is in execution. This has been achieved by keeping a small value of the threshold T_{MPT}.

In Fig. 8, some of the video frames extracted from the results obtained using MPT algorithm are shown. It is evident that the leg of the baby is being successfully tracked using the proposed MPT algorithm which was not possible using the single path algorithm.

2.9 Analysis of the tracking algorithm

In this section, computational performance of the proposed tracking algorithm is presented. An estimation of the approximate number of nodes being processed during a complete execution of the algorithm has been carried out in this section. Let a video of the lateral tilting test contains N frames and the level of MPT is assumed to be $L_{\max}$. In the worst case, the tracking algorithm may run in MPT mode during the entire duration of the video. Thus, there will be a maximum of $\lceil N/L_{\max} \rceil$ ($N \gg L_{\max}$) number of calls to the MPT function assuming that each MPT is explored up to its fullest level $L_{\max}$. In such a scenario, total umber of nodes being processed is estimated to be $\lceil N/L_{\max} \rceil \frac{K^{L_{\max}}-1}{K-1}$, which runs in $O(NK^{L_{\max}})$ time where K is the number of paths being explored inside MPT. If the algorithm runs in single path mode for the entire duration of the video, N number of nodes need to be processed.

3 Results and assessment

Performance of the proposed tracking algorithm has been evaluated with 42 sample videos of lateral tilting examinations that were recorded at the neurodevelopment clinic of SSKM Hospital, Kolkata during March 2010–September 2011. A desktop computer powered by Intel Dual Core processor @2.13 GHz with 3 GB of DDR2 memory running

Linux 2.6 kernel and gcc version 4.1.2 was used during the trials. The algorithm has been implemented using the Intel open source computer vision (OpenCV) library.[1] The first phase of the rotation was carefully analyzed with the recorded data using the proposed MPT algorithm with Kalman filter-based prediction having state transition matrix as in (6).

Though the proposed multi-path-based tracking algorithm is computationally expensive than the single path algorithm; however, it was observed that the performance gain using multi-path tracking algorithm over single path is significant. As shown in Figs. 5 and 8, performance of the single path algorithm is not reliable, whereas the multi-path algorithm does not fail to track feature points in the present context. Since the videos are analyzed offline, it was felt that performance gain be given more importance in spite of slight increase in computational overhead. In fact, it was observed that for a video of one minute duration, processing time saved by the single path tracking algorithm is in the order of a few seconds only.

Typical trajectories of various body parts including head, torso, and limbs obtained following the MPT algorithm with Kalman filter-based prediction are shown in Fig. 9. Tracking results superimposed on some of the frames of the sample video are shown in Fig. 10. In this figure, tracked body parts are highlighted using white circles. It is evident from Fig. 10 that the proposed MPT-based algorithm successfully tracks the body parts during tilting. The trajectories of various body parts are found to be elliptical. It may be observed from Fig. 9 that the trajectories deviate from the usual elliptical paths once the tilting phase is over. This is due to the movement of head and limb after the rotation. However, we may observe occasional deviation of the limbs from the usual path due to high degrees of freedom of the limbs. For example, the trajectory of the left leg shown in the figure is significantly different from other trajectories. However, according to the domain experts, this can be ignored in the present context since an observer should ideally look for involuntary movements of the body parts only after the rotation is over. Therefore, an algorithm as discussed below can be used to detect the end point of the rotation without any difficulty caused by such noises.

Break point detection algorithm:

- Step 1: Let $V = \langle v_1, v_2, \ldots, v_n \rangle$ be the trajectory of a body part that has been tracked using the proposed MPT algorithm, where n represents the total number of points in the trajectory. Assume, W_B be the size of an window that is used in the break point detection algorithm.
- Step 2: Get the direction of movement according to W_B successive points on the trajectory V with respect to the epicenter of the ellipse detected in the first phase.

[1] http://sourceforge.net/projects/opencvlibrary/.

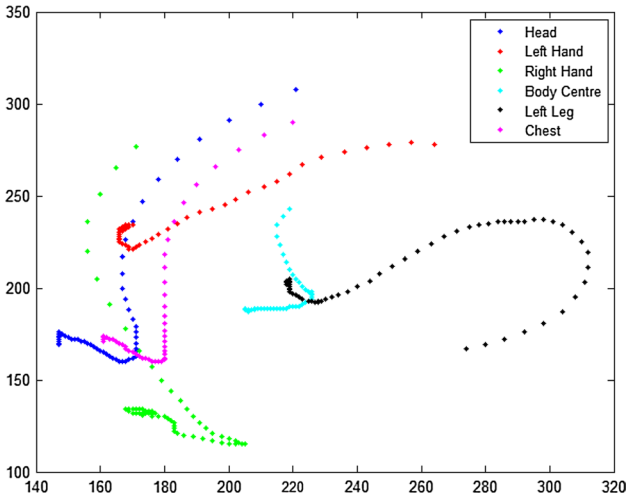

Fig. 9 Graphical representation of the trajectories of various body parts of a baby undergoing the lateral tilting test

- Step 3: If the direction of the movement is reversed in two successive non-overlapping windows, mid point between these windows can be assumed to be the break point. Else, continue Step 2 to process the next window.

The above mentioned break point detection algorithm successfully identifies the end of the first phase and beginning of the reflex reactions. To make the break point detection algorithm robust, it is supported by multiple evidence using all of the 8 feature points mentioned in the 2D geometrical model shown in Fig. 2. A majority voting is taken while estimating the final position of the break point. Next, after detection of the break point (end of tilting), translational motion of the body parts are estimated and the angle as mentioned in Table 1 are measured to assess the test and assign a score accordingly.

Earlier, it has been described that there exists a significant inter-observer (0.089) as well as intra-observer (0.095) variances [3] in the conventional scoring method. As there is a chance of subjectiveness and bias in the visual observation-based assessment method, we have compared the results of the proposed computer-based assessment with physician's evaluation to support the utility of the proposed computer-aided approach. Comparisons have been carried out using well known measures, namely precision and recall.

The comparison has been carried out as follows. In this quest, first, comparisons are done independently with physician's assessments and then a majority voting-based analysis is presented. The precision and recall values are computed from the observed true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values. These four measures are determined for each class, namely Score 0, Score 1, Score 2, and Score 3.

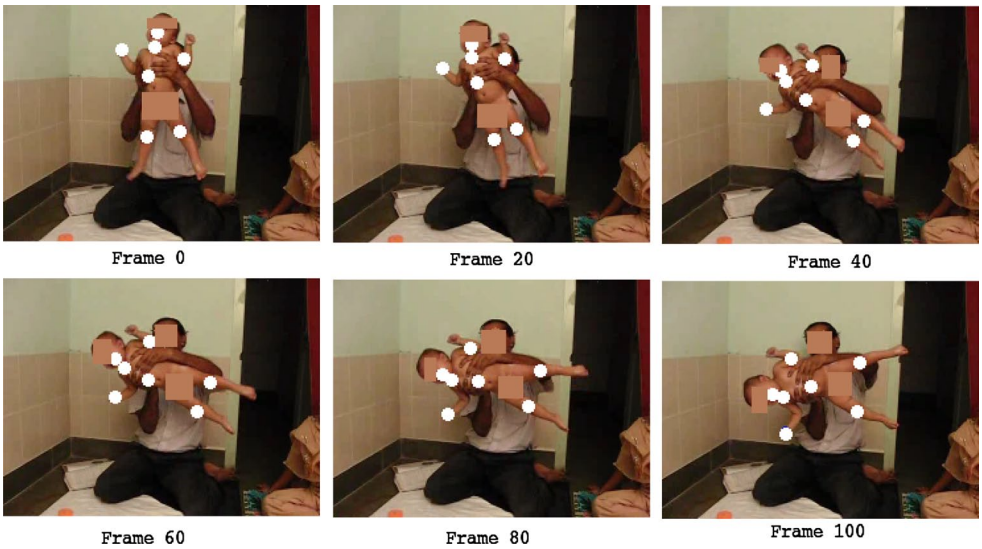

Fig. 10 Some of the image frames extracted from the video recording of lateral tilting examination

Table 2 Comparison of performance between computer-based assessment and physician's measurement of lateral tilting test carried out on 42 babies

	Score 0	Score 1	Score 2	Score 3
	TP 2	TP 1	TP 2	TP 22
Physician 1	FN 0	FN 2	FN 4	FN 12
Observation 1	FP 0	FP 3	FP 8	FP 4
	TN 40	TN 37	TN 31	TN 8
	Sensitivity 100 %	Sensitivity 33 %	Sensitivity 33 %	Sensitivity 65 %
	Specificity 100 %	Specificity 92.5 %	Specificity 79.5 %	Specificity 67 %
	TP 2	TP 2	TP 4	TP 23
Physician 1	FN 0	FN 0	FN 4	FN 6
Observation 2	FP 0	FP 2	FP 6	FP 2
	TN 40	TN 38	TN 28	TN 10
	Sensitivity 100 %	Sensitivity 100 %	Sensitivity 50 %	Sensitivity 79.3 %
	Specificity 100 %	Specificity 95 %	Specificity 82.4 %	Specificity 83.3 %
	TP 2	TP 1	TP 3	TP 24
Physician 2	FN 0	FN 0	FN 5	FN 7
Observation 1	FP 0	FP 3	FP 7	FP 3
	TN 40	TN 38	TN 29	TN 9
	Sensitivity 100 %	Sensitivity 100 %	Sensitivity 37.5 %	Sensitivity 77 %
	Specificity 100 %	Specificity 92.6 %	Specificity 80 %	Specificity 75 %
	TP 2	TP 2	TP 5	TP 24
Combined	FN 0	FN 2	FN 7	FN 12
	FP 0	FP 2	FP 4	FP 2
	TN 40	TN 36	TN 26	TN 4
	Sensitivity 100 %	Sensitivity 50 %	Sensitivity 41.6 %	Sensitivity 67 %
	Specificity 100 %	Specificity 94.7 %	Specificity 83.8 %	Specificity 67 %

True positive (TP) for a class, say Score 0, is defined as the total number of infants that have been rightly classified with a score of 0 using the proposed computer-aided method and the physician agrees with this classification. Similarly, true negative (TN) for a class is defined as the total number of infants that have been rightly classified as a non-member of this particular class, say Score 0, and belong to any of the remaining classes (e.g., Score 1, Score 2, or Score 3). False positive (FP) is defined as the number of infants who according to a physician does not belong to a particular class, however, graded falsely into that class by the proposed method. False negative (FN)

represents the total number of infants who should have been graded to a particular class according to the physician, however, graded wrongly to other classes. Table 2 represents the sensitivity ($\frac{TP}{TP+FN}$) and specificity ($\frac{TN}{TN+FP}$) of various comparisons considering more than one physicians evaluation.

In the combined assessment, if anyone of the scores given by the physicians matched with the computer-based measurement, the sample was assumed as true positive. It may be observed that, 27 out of 42 total samples were classified rightly by Physician 1 during his/her first observation while at 31 instances, they were found to be rightly classified by the same physician when the second opinion was taken. Therefore, out of a total of 84 instances, the computer-based assessment agree with Physician 1 in 58 cases with an accuracy of 69 %. Similarly, an accuracy of 71 % was recorded in case of Physician 2.

4 Conclusion and future work

In this paper, a new multi-path trajectory analysis-based video object tracking algorithm has been proposed. We have adopted a greedy approach for selecting multiple candidate feature points and then a prediction of correct block is done using Kalman filter. A single path-based tracking approach has been generalized to explore multiple trajectories which are refined using the filter. The proposed algorithm has been successfully used to analyze the video recordings of the lateral tilting test. It has been demonstrated in this paper that the proposed methodology can assist physicians of neurodevelopment clinics in the decision making process of HINE. However, an extensive field trial is necessary to test the robustness of the algorithm in long run. It is believed that, tracking of body parts of infants during other high speed movements-related examinations of the HINE set can be done using the proposed tracking algorithm. This may be applied for tracking fast moving objects in other context.

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 Dr. Arunava Biswas and Asish Ghosh of the neurodevelopment clinic of the SSKM Hospital for their 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. Face and other visible organs of the subjects as well as the examiners have been blurred in the images for maintaining secrecy.

Conflict of interest Authors have no conflict of interest with anyone linked with this work.

References

1. Calvagno G, Rinaldo R, Sbaiz L (1998) Three-dimensional motion estimation of objects for video coding. IEEE J Sel Areas Commun 16(1):86–97
2. Cox I, Hingorani S (1996) An efficient implementation of Reid's multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking. IEEE Trans Pattern Anal Mach Intell 18(2):138–150
3. Dogra D (2012) Algorithms for video assisted analysis of infant neurological examinations. PhD Thesis, Indian Institute of Technology Kharagpur
4. Dogra D, Nandam K, Majumdar A, Sural S, Mukhopadhyay J, Majumdar B, Mukherjee S, Singh A (2011) A tool for automatic Hammersmith infant neurological examination. Int J E-Health Med Commun 2(2):1–13
5. Dogra D, Majumdar A, Sural S, Mukherjee J, Mukherjee S, Singh A (2012) Toward automating Hammersmith pulled-to-sit examination of infants using feature point-based video object tracking. IEEE Trans Neural Syst Rehabil Eng 20(1):38–47
6. Dogra D, Majumdar A, Sural S, Mukherjee J, Mukherjee S, Singh A (2011) Automatic adductors angle measurement for neurological assessment of post-neonatal infants during follow up. In: Pattern recognition and machine intelligence, international conference on, pp 160–166
7. Dogra D, Nandam K, Majumdar A, Sural S, Mukhopadhyay J, Majumdar B, Mukherjee S, Singh A (2010) A user friendly implementation for efficiently conducting Hammersmith infant neurological examination. In: E-Health networking, application and services, IEEE international conference on, July, pp 374–378
8. Dubowitz L, Dubowitz V, Mercuri E (2000) The neurological assessment of the preterm and full term infant, vol 9. Clinics in Developmental Medicine. Heinemann, London
9. Gao T, Li G, Lian S, Zhang J (2011) Tracking video objects with feature points-based particle filtering. Multimed Tools Appl 58(1):1–21
10. Han B, Roberts W, Wu D, Li J (2007) Robust feature-based object tracking. In: Algorithms for synthetic aperture radar imagery, proceedings of SPIE conference on, vol 6568
11. Han B, Roberts W, Wu D, Li J (2007) Robust feature-based object tracking. In: Society of photo-optical instrumentation engineers conference series, vol 6568, April
12. Kalman R (1960) A new approach to linear filtering and prediction problems. Trans ASME J Basic Eng 82:35–45
13. Kasprzak W (2000) Adaptive methods of moving car detection in monocular image sequences. Mach Graph Vis 9(1/2):167–185
14. Lee M, Nevatia R (2009) Human pose tracking in monocular sequence using multilevel structured models. IEEE Trans Pattern Anal Mach Intell 31(1):27–38
15. Liu Q, Sclabassi R, Sun M (2004) Change detection in epilepsy monitoring video based on Markov random field theory. In: Intelligent signal processing and communication systems, proceedings of international symposium on, November, pp 63–66
16. Lowe D (1999) Object recognition from local scale-invariant features. In: Computer vision, proceedings of the seventh IEEE international conference on, vol 2, pp 1150–1157
17. Miao Q, Wang G, Lin X, Wang Y, Shi C, Liao C (2010) Scale and rotation invariant feature-based object tracking via modified on-line boosting. In: Image processing, IEEE international conference on, September, pp 3929–3932
18. Nishida Y, Motomura Y, Kitamura K, Mizoguchi H (2004) Infant behavior simulation based on an environmental model and a developmental behavior model. In: Systems, man and cybernetics, proceedings of IEEE international conference on, vol 2, October, pp 1555–1560

19. Noble J, Boukerroui D (2006) Ultrasound image segmentation: a survey. IEEE Trans Med Imaging 25(8):987–1010
20. Park S, Aggarwal J (2002) Segmentation and tracking of interacting human body parts under occlusion and shadowing. In: Motion and video computing, proceedings of workshop on, December, pp 105–111
21. Romeo D, Guzzetta A, Scoto M, Cioni M, Patusi P, Mazzone D, Romeo M (2008) Early neurologic assessment in preterm-infants: integration of traditional neurologic examination and observation of general movements. Eur J Paediatr Neurol 12(3):183–189
22. Romeo D, Cioni M, Scoto M, Mazzone L, Palermo F, Romeo M (2008) 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):24–31
23. Roy A, Sural S, Mukherjee J, Majumdar A (2008) State-based modeling and object extraction from echocardiogram video. IEEE Trans Inf Technol Biomed 12(3):366–376
24. Singh S, Hsiao H (2003) Infant telemonitoring system. In: Engineering in medicine and biology society, proceedings of the 25th annual international conference of the IEEE, vol 2, September, pp 1354–1357
25. Song H, Shen M (2011) Target tracking algorithm based on optical flow method using corner detection. Multimed Tools Appl 52(1):121–131
26. Tran T, Lee P, Pham V, Shyu K (2008) MRI image segmentation based on fast global minimization of snake model. In: Control, automation, robotics and vision, proceedings of the international conference on, December, pp 1769–1772
27. Vohr B, Wright L, Mele A, Veter J, Steichen J, Simon N, Wilson D, Broyles S, Bauer C, Delaney-Black V, Yolton K, Fleisher B, Papile L, Kaplan M (2000) 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):1216–1226
28. Wang J, Chen X, Gao W (2005) Online selecting discriminative tracking features using particle filter. In: Computer vision and pattern recognition, IEEE computer society conference on, vol 2, June, pp 1037–1042
29. Yilmaz A, Javed O, Shah M (2006) Object tracking: a survey. ACM Comput Surv 38(4):1–45
30. Zhang R, Vogler C, Metaxas D (2007) Human gait recognition at sagittal plane. Image Vis Comput 25(3):321–330