A video/IMU hybrid system for movement estimation in infants

Machireddy, Archana; van Santen, Jan; Wilson, Jenny L; Myers, Julianne; Hadders-Algra, Mijna; Xubo Song

Published in:

DOI:
10.1109/EMBC.2017.8036928

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2017

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
A Video/IMU Hybrid System for Movement Estimation in Infants

Archana Machireddy, Jan van Santen, Jenny L. Wilson, Julianne Myers, Mijna Hadders-Algra, Xubo Song

Abstract— Cerebral palsy is a non-progressive neurological disorder occurring in early childhood affecting body movement and muscle control. Early identification can help improve outcome through therapy-based interventions. Absence of so-called “fidgety movements” is a strong predictor of cerebral palsy. Currently, infant limb movements captured through either video cameras or accelerometers are analyzed to identify fidgety movements. However both modalities have their limitations. Video cameras do not have the high temporal resolution needed to capture subtle movements. Accelerometers have low spatial resolution and capture only relative movement. In order to overcome these limitations, we have developed a system to combine measurements from both camera and sensors to estimate the true underlying motion using extended Kalman filter. The estimated motion achieved 84% classification accuracy in identifying fidgety movements using Support Vector Machine.

I. INTRODUCTION

It is estimated that around 15 million babies are born preterm (before 37 weeks of gestation) worldwide every year [1]. Preterm infants are at increased risk of developing neurological and motor impairments [2]. The most common motor impairment affecting preterm infants is Cerebral Palsy (CP) and is found in 4-20% of preterm infants based on their gestational age [3]. Cerebral palsy is a non-progressive disorder causing functional impairment of the brain. The neurological damage leads to complex dysfunctions affecting movement, motor skill and muscle co-ordination. Although the primary damage cannot be repaired, identifying CP at an early stage enables early intervention, which can help minimize resultant impairments [4].

Currently, the diagnosis of CP in infants is based on the assessment of their spontaneous “general movements”, i.e., movements made while lying in supine position on a flat surface, not surrounded by any objects that obstruct movement or attract attention. The development of movements in early infancy is a good predictor of movement and cognitive performance later on [5]. Prechtl and colleagues [6] studied the development of spontaneous movements in infants, and came up with Prechtl’s General Movement Assessment (GMA), which can predict CP with sensitivity of 98% [7]. Notably, the absence of fidgety movements (FM) in infants between 2 to 4 months has shown to be a strong marker in identifying infants who will develop CP [8]. Fidgety movements are defined as small circular motions of moderate speed with variable acceleration [6].

Although GMA can predict CP with high accuracy, it is not widely adopted clinically, as it requires clinicians to be trained specifically in the Prechtl’s GMA method and judgments are subjective [8]. This led to the development of automated methods for analyzing motion in infants. An infant’s limb movements can be captured by using a video camera or wearable sensors such as accelerometers, and analyzed to identify different movements [9]. Video based movement analysis gives high spatial resolution and is easy to interpret, but can be hampered by occlusion, and generally has low temporal resolution unless expensive special-purpose high frame rate cameras are used. To predict fidgety movements it is necessary to capture subtle changes in movements of the limbs, which might not be possible using video based methods. In contrast, accelerometer based methods have high temporal resolution, which can be helpful to track subtle changes, but they lack spatial information [9].

In this work we developed a hybrid system, which combined measurements from a basic camera and wearable sensors to benefits from the superior spatial and temporal resolution capabilities of either modality. Specifically, we used 9-dimensional inertial measurement units (IMUs), containing 3D accelerometer, 3D magnetometer, and 3D gyroscope, unlike other studies where only one of these sensors was used [9]. The true 3D motion was estimated using an extended Kalman filter (EKF), which combined measurements from video and IMU. The estimated motion was used to classify fidgety movements from non-fidgety movements using Support Vector Machine (SVM).

II. RELATED WORK

A. Video-based assessment

Meinecke et al. performed 3D motion analysis by placing reflective markers on infant and capturing motion using seven infrared cameras [10]. They extracted 53 quantitative parameters among which they found an optimal combination of eight parameters by cluster analysis. These optimal eight parameters were used as features for quadratic discriminant analysis to obtain an overall CP detection rate of 73%. But such 3D motion capture systems are costly, hard to set-up, and have high computational complexity, thus limiting their clinical use. Kanumera et al. also used reflective markers, but used a digital video camera to capture motion. They...
examined different features extracted from the 2D-positional data from the video. They found that the movements of infants who develop CP were jerkier than those of normal infants [11]. As only the 2D-positional data was considered, movement perpendicular to the camera is completely overlooked.

Adde et al. developed a general movements toolbox to detect fidgety movements in video recordings [12]. They used the 2D representation of movement over time, called the motiongram, to extract eight quantitative features. They achieved a sensitivity of 81.5% in classifying fidgety movements. Stahl et al. extracted motion information by applying optical flow [13]. Features were extracted using wavelet analysis and classified using SVM to achieve 93% accuracy in identifying fidgety movements. But these studies used marker-less tracking, which could be less accurate as subtle differences in moving patterns cannot be detected due to low temporal resolution of camera. The advantage of using standard video cameras is the reduced cost of the system and the set-up effort, enabling applications beyond research setting, to regular clinics as well as allowing for continuous analysis even at home.

B. Sensor-based assessment

Gravem et al. used accelerometers to monitor movements in ten preterm infants to identify cramped synchronous general movements (CSGM) [14]. Their statistical models were able to predict CSGMs with 70-90% accuracy. No attempts were made to predict CP and the age of the infants studied (30-43 weeks) was too old for predicting CP from general movements [15].

Heinze et al. extracted 32 features based on velocity and acceleration extracted from the accelerometer data for 19 healthy and 4 affected subjects [16]. They used decision trees to obtain a detection rate between 88-92%. Philipp et al. used a magnetic tracking system to record movements [17, 18]. They extracted three kinematic features out of which repetitive movement in the upper limbs proved to be a good predictor of CP. But the accelerometer and magnetic tracking system used by above two studies were wired and large in size, posing significant practical problems.

C. Hybrid systems

Berge et al. proposed a software tool called ENIGMA (enhanced interactive general movement assessment) which helped visualize movement patterns seen in video [19]. Here the motion is captured using a video camera and by measuring position and orientation of a sensor with respect to the pulses of magnetic field. They proposed a periodicity feature for detecting fidgety movement, but did not provide any quantitative analysis on how well this feature performed.

III. METHOD

A. Participants

Twenty infants between the ages of 2-4 months were recruited for this study. Written consent was obtained from all parents. Infants were placed in the supine position on a mattress with sensors attached, and a stationary video camera was placed above, which captured movement in the horizontal plane. A video was recorded for 30 minutes when the infant was in an active wakeful state.

Figure 1. (a) Baby with the Shimmer sensors and color patches (b) detected color patches

B. Sensor Recording

We used five Shimmer3 sensors from Shimmer Inc. [20], attached to left leg, left hand, right leg, right hand and chest using soft wrist and ankle bands, and a vest as shown in Fig. 1. Shimmer3 sensors have a 3D accelerometer, a 3D gyroscope and a 3D magnetometer. A sampling rate of 256Hz was used. To calibrate the sensor, the constant bias in the accelerometer and gyroscope were determined beforehand from static recordings and subtracted from the measurements.

C. Camera Recording

We used a down-pointing video camera (Panasonic HC-V550), which was calibrated to obtain intrinsic parameters (focal length, scaling and skewing factors along the horizontal and vertical axis of the image plane) and extrinsic parameters (rotation and translation of the camera coordinates relative to the world coordinates). To avoid the difficulty of calibrating camera arrays and reducing distortion to the infants, we used only one camera, which captures only the 2D projection of true 3D motion. We used a fiducial-based motion tracking approach. We have color patches sewn on the wrist and ankle bands holding the Shimmer3 sensors as shown in Fig. 1. The locations of these patches on the image plane can be detected accurately since each patch has a unique and predetermined color. We placed five patches of different colors, one each on four sides of the wrist and one on top of the mitten so that we have at least one patch visible at all times, irrespective of the orientation of the hand. Since the relative positions of the color patches to one another are known, the position of the color patch on the shimmer can be estimated, even when it is occluded. To detect the patches we used the HSV (hue, saturation, value) color space. The main advantage of HSV over RGB color space is that it separates color from intensity information. We set thresholds, using the exact range of hue, saturation and value measures representing each color patch to detect them. The true 2D positions of the patches were estimated by back projecting the point on the image plane through the camera calibration matrices. The color patches were also used to obtain an estimate of depth (z dimension). As the true size of the color patch (w) and the focal length of the camera (f) were known, we used the triangular similarity to get its distance from the camera using the formula $D = (w * f) / p$, where p is the width of the patch on the image in pixels [21]. Therefore, this provided the 3D spatial position of the colored patch.

The sampling rate for the video signal was 60 fps, while for shimmer it was 256 Hz. To synchronize these two modalities, a buzzer connected to an additional Shimmer was
used. It simultaneously produced a 4175 Hz audio signal and a voltage spike in the Shimmer signal. As all Shimmers were synchronized, the timing of voltage spike was looked up on the other Shimmers. The Goertzel algorithm [22] was used to detect the 4175Hz audio signal from the video, which was then aligned with all the Shimmers. To estimate the 3D motion, the 3D position estimates from the camera were fused with the IMU signals, using the EKF framework described below.

D. Combining measurements using extended Kalman filter

The EKF propagates an estimate of the system state $x$ (e.g. position, orientation), given a sequence of observations $y$ (e.g. Shimmer output, image features) using the equations $x_{k+1} = f(x_k) + \eta_k$ and $y_k = h(x_k) + \epsilon_k$, where $f$ is the system model, $h$ the observation model, and $\eta$ and $\epsilon$ are the system and observation noise, respectively. We defined the state vector as $x = [p \ v \ a \ q \ w]^T$, where $p$, $v$ and $a$ are the 3D position, velocity and acceleration, $q$ is the 4 x 1 quaternion describing rotation from global frame to sensor frame, and $w$ is the angular velocity. The state estimates were propagated using the state transition model and noise covariance matrix ($Q_k$) as shown in Fig. 2. This is a simple model similar to previous work on modeling human motion [23], assuming constant acceleration and angular velocity. The actual variation was modeled by zero-mean white Gaussian noise with covariance matrices $\Sigma_a = \sigma_a^2I$ and $\Sigma_w = \sigma_w^2I$, respectively.

The observation vector was defined as $y = [p \ a \ w \ m]^T$, where $p$, $a$, $w$ and $m$ are the measurements from accelerometer, gyroscope and magnetometer, respectively. The observation equations and the observation noise covariance matrix ($N_k$) are shown in Fig. 2. As the two modalities were sampled at different rates, a multi-rate filtering strategy was used. The exact time instances of the visual and shimer measurements were known, therefore the time difference between successive measurements (dt) was known. This time difference was used in the state equations, and the rows of the observation vector, for the modality missing the measurement at that instant, were set to zero.

The reason to adopt the EKF approach, which is based on the first-order linearization using Taylor expansion, is the nonlinear nature of the measurement equations (due to the presence of rotation matrix $R$). It can be seen from Fig. 2 that the observed acceleration ($a_o$) is the sum of the true body acceleration ($a_b$) and the gravitational acceleration ($g$) as perceived in the body frame. As the magnetic field in a room can be distorted due to the presence of other metal structures, a distortion compensation method proposed by Madgwick et al. [24] is used, wherein the earth’s magnetic field is represented as $b = [\sqrt{h_x^2 + h_y^2} \ 0 \ h_z]$, where $h = Rm_o$, so that it is has the same inclination as the measurement. This hybrid system provided an estimate of the 3D position, velocity, acceleration, orientation and angular velocity of the infant’s limbs with high spatial and temporal resolution. As fidgety movements are characterized by their velocity and acceleration, we used the estimated 3D velocity, acceleration, angular velocity and their magnitudes to predict fidgety movement.

IV. RESULTS

A. 3D motion estimation using simulated data

To evaluate the accuracy of the EKF model in predicting the position and orientation and in particular the prediction of depth, we made a model of human arm using plywood, and simulated circular motion by rotating it using a drill. Knowing the length of the arm, exact position and angle of the drill, the ground truth movement was determined. A Shimmer sensor was placed on the plywood-arm with a color marker and the motion was recorded using a camera. This motion was estimated using EKF as explained above. From Figs 3(a) and 3(b) it can be seen that the EKF framework is able to correctly predict the position and orientation of the plywood-arm. We simulated other movements such as the 3D spiral motion and the model was able to estimate the positions accurately as seen from Figs 3(c).

B. 3D motion estimation in infants

The video of an infant was analyzed by an expert to mark time intervals with fidgety movements and non-fidgety
movements. The motions of different limbs were estimated using the hybrid system. The estimated velocity, acceleration, angular velocity and their magnitude values were chosen to form a 12 dimensional feature at each time instant. Equal lengths of fidgety and non-fidgety movement segments were chosen to form a balanced dataset. A 10-fold classification using SVM on data from all limbs gave 84% accuracy in classifying fidgety from non-fidgety movements. A 10-fold classification on data from a single limb (i.e. train and test data from same limb) gave an accuracy of 90%, while using train and test data from different limbs gave 70% classification accuracy.

V. CONCLUSION

A novel method combining video and IMU sensor inputs to estimate 3D infant body movements is presented. A new approach to estimate depth by using markers captured by the video is demonstrated. Accelerometer, gyroscope and magnetometer features are combined with the positional information from video using EKF to derive the 3D position and orientation of the limbs of infant. The proposed method is shown to be able to accurately estimate 3D motion.

Based on the estimated motion, fidgety movements were classified with over 84% accuracy. The results in detecting fidgety movements are promising on the small dataset used, but need to be further assessed on a larger population. A dynamic model characteristic of infant’s limb movements can give better features and deeper insights into the behavior of the fidgety movement. With the proposed model’s estimates of position, velocity, acceleration and orientation, improved dynamic features can be computed to further help capture the temporal behavior of fidgety movement and aid in increasing the classification accuracy. The applicability of fidgety movements in predicting CP needs to be evaluated in further studies by recruiting high-risk population and carrying out a long-term neurological study.

ACKNOWLEDGMENT

This research is made possible by a generous gift from Henry T. Swigert. The content is solely the responsibility of the authors and does not necessarily represent the views of the donor.

REFERENCES