This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2019.2916163, IEEE Sensors Journal

Automatic Synchronization of Markerless Video and Wearable Sensors for Walking Assessment

Yi Chiew Han, *Student Member, IEEE*, Kiing Ing Wong, *Member, IEEE*, and Iain Murray, *Senior Member, IEEE*

***Abstract*— As walking assessment is commonly done through visual inspection, it is beneficial to make joint angle information available to the clinicians for better assessment. The main challenge is that the video camera and inertial sensor are usually two separate systems, and the recordings are hard to be initialized at the same time manually. This creates a problem that the inertial sensor data is not temporally synchronized with the video camera. This paper proposes a method to synchronize the video and sensor data by detecting and matching the maximum backward swings of the leg. The proposed method is validated by blinking LED and transmitting LED flag to the computer at the same time. The synchronization error of the proposed method is low at about 0±2 frames compared to the validation method.**

***Index Terms*— Angle, inertial sensor, synchronization, video.**

I. Introduction

IN MEDICAL FIELD, walking assessment is performed by clinicians to provide optimal care and treatments for patients [1]. The fundamental part of walking assessment is the estimation of joint position and orientation [2]. At current stage, walking assessment is commonly done through visual inspection which strongly depends on the experience of clinicians [3]. It is beneficial for clinicians to have access to joint angle information for better walking inspection.

Optical motion capture system such as Vicon [4] and OptiTrack [5] can be used to estimate the joint position and orientation accurately. However, such system is expensive and non-portable. An Inertial Measurement Unit (IMU), on the other hand, is an affordable and portable electronic device that consists of accelerometers and gyroscopes that can be used to estimate joint orientations [6].

However, IMU cannot capture video of the joint movement for visual inspection. A video camera can be used together with IMU such that the video and IMU data are recorded simultaneously. Some researchers used IMUs and video cameras for motion tracking [7], localization [8], and video stabilization [9]. To reduce cost, a smartphone camera is used in this research.

The use of synchronization hardware module is common, so that the video and IMU data can be recorded at the same time. However, when the video camera and IMU are two separate systems and the system clocks are not accessible, the manual recordings of video and IMU data are hard to be initialized at the same time due to human and software delay [10]. Therefore, there is a need to perform temporal synchronization for the video and IMU data through some signal processing.

We proposed to temporally synchronize video and IMU data of a person walking on a flat surface by detecting and matching the maximum backward swing of the leg from video and IMU. The proposed method is validated by blinking LED and sending LED flag to computer (PC) at the same time.

The remaining of this paper is structured as follows. Related works are stated in Section II. The methods to collect data are stated in Section III. Section IV describes our proposed method to synchronize video and IMU data of a person walking on a flat surface. Section V is the results and discussion. Conclusion and future direction are stated in Section VI.

II. Related Works

In this paper, we proposed to match the maximum backward swing of leg detected from video and inertial sensor for synchronization. There are several possible methods to detect the maximum backward swing of leg from the video. Vicon motion capture system [4] is the gold standard in estimating the joint position and orientation, but this system is expensive and requires multiple cameras. Meribout *et al.* [11] used a parallel hardware architecture to support a parallel Hough transform algorithm to recognize the shape of an object. The method can be extended to detect the shape of any kind of objects including legs, therefore the current angle of legs can be estimated. Other edge detection method such as that proposed by Hu et al. [12] can also be used to detect the leg. Zhang *et al.* [13] proposed a joint gait–pose manifold-based visual gait generative model to estimate 3D gait kinematics from a single video camera. Although the proposed method is accurate enough for some applications, the estimated joint orientations still deviate greatly from the ground-truth at some video frames. On the other hand, existing inertial sensor-based orientation estimation algorithms have achieved higher accuracies [14] [15].

There are several advantages of synchronizing video and inertial sensor data. Chen *et al.* [16] recognized 27 different human motions such as walking and arm-swinging using a

Y. C. Han, and K. I. Wong are with the Department of Electrical and Computer Engineering, Curtin University Malaysia (e-mail: yc.han@postgrad.curtin.edu.my, wong.kiing.ing@curtin.edu.my).

I. Murray is with the School of Electrical Engineering, Computing and Mathematical Sciences, Curtin University Australia (e-mail: I.Murray@curtin.edu.au).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2019.2916163, IEEE Sensors Journal

Microsoft Kinect camera and an IMU. The accuracy of activity recognition when camera and IMU were fused was about 10% higher than that of using camera or IMU alone. Farnoosh *et al.* [17] fused inertial data and video recorded from a smartphone for indoor navigation. The inertial sensor was used to estimate the smartphone orientation, and the navigation accuracy was improved compared to navigation without orientation estimation. Jatesiktat *et al.* [18] fused a Kinect's depth camera with two IMUs worn on the wrists to improve the accuracy of the upper-body joint tracking. Validated against the gold standard Vicon system, the authors have successfully improved the Kinect's skeleton tracking by 20%.

Bae *et al.* [19] designed a synchronization hardware module which connects gyroscope and camera to synchronize them. In cases where the video camera and inertial sensors are two separate systems without a synchronization hardware module, the recordings are hard to be initialized at the same time manually. Signal processing needs to be done to synchronize the video camera and inertial sensor data. Cippitelli *et al.* [20] synchronized depth cameras connected to a computer and a wireless IMU connected via Bluetooth to the same computer. The transmission time delay between the cameras and the computer was estimated by blinking LEDs controlled by an Arduino board connected to the same computer. Plotz *et al.* [21] synchronized camera and accelerometer data using cross-correlation based time-delay estimation. The horizontal and vertical hand gestures have low average error of 0.5 and 2 frames, respectively. However, circular and wave-like gestures have error more than 10 and 20 frames, respectively.

Synchronization methods for multiple videos are also reviewed. Lin *et al.* [22] synchronized two videos captured at different angles. The authors first detected the upper body of the subjects, then compared the brightness of the upper body for correlation-matching between the videos. The average synchronization error was within 1 frame. Duong *et al.* [23] synchronized multiple versions of the same movie by matching the audio tracks.

Ryu *et al.* [24] used a position sensitive detector (PSD) camera module to identify the positions of markers attached to a moving object. Each marker consists of a radio frequency transmitter and an infrared LED. The LED blinked and at the same time transmitted a command that consists of the marker's identification number to the PSD to distinguish the identity of each marker.

Ofli *et al.* [25] introduced Berkeley Multimodal Human Action Database which consists of temporally synchronized video, audio and accelerometer data of people performing activities such as jumping and sitting. The authors mentioned that the video, audio, and accelerometer data were recorded simultaneously.

Overall, synchronizing video and inertial sensor data require signal processing as the recordings of video and inertial sensor data are hard to be initialized at the same time due to human and software delay. Besides, the video and inertial sensor may be sampled at different rate.

III. Data Collection

A. Inertial Measurement Unit (IMU)

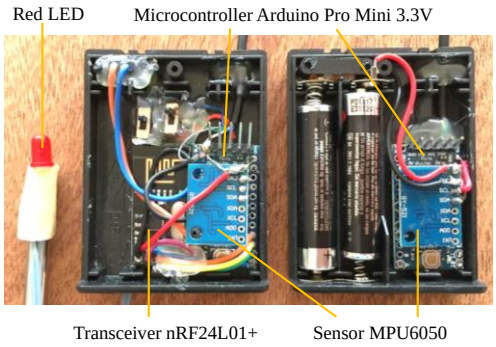


Fig. 1. IMU and LED.

Two inertial measurement units (IMUs), as shown in Fig. 1, were constructed. Each IMU consists of a microcontroller Arduino Pro Mini 3.3V and a sensor MPU6050. MPU6050 has a tri-axis accelerometer with ±2g range and a tri-axis gyroscope with ±250º/s range. Both sensors are time-synchronized and sampled at 100Hz with 16-bit resolution [26].

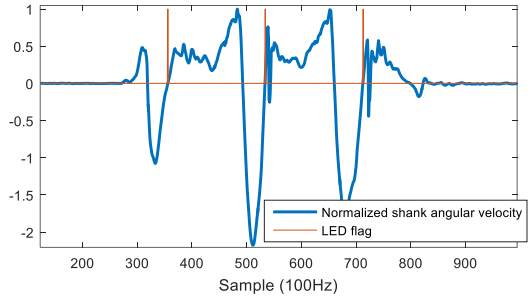


Fig. 2. Blinking of LED for validation of proposed method.

A red LED is also connected to the IMU through wire. The LED blinks for 10ms when there is a positive zero crossing of the shank's angular velocity and the either one of the previous 10 shank's angular velocities is lower than -100°/s, as shown in Fig. 2. At the same time of the LED blinks, an LED flag = 1 is encoded in the IMU wireless data transmission to the PC.

Sensor readings from the foot's IMU are transmitted to the shank's IMU through wires, and then the shank's IMU transmits all raw sensor data and the LED flag to PC through wireless transceivers nRF24L01+.

The IMUs and the LED are powered by a total of 2 AAA batteries.

B. Video camera

The video camera used in this research was the front camera of an iPhone 6 Plus, which records 720p HD video with a resolution of 1280 × 720 pixels at a frame rate of 30fps [27].

C. Experimental setup

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2019.2916163, IEEE Sensors Journal

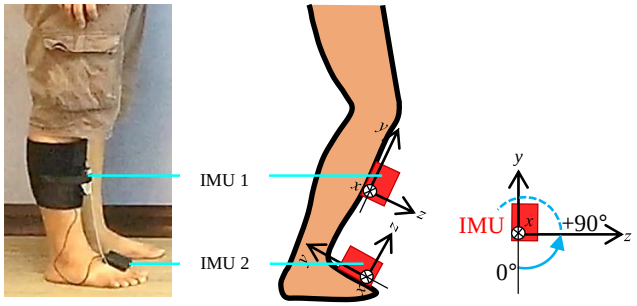


Fig. 3. IMU placement.

Fig. 3 show that the IMUs were strapped in front of the right shank using Velcro straps, and on top of the right foot using double-sided tape without any skin penetration. The red LED was placed at the bottom left of the video.

10 healthy adults (male: 7; female: 3; age: 21-49; height: 151-182cm) participated in the data collection. Each person was asked to walk self-pace on a flat surface for about 3 meters. In the first 5 trials, the participants were asked to start walking with their right legs. In the next 5 trials, the participants were asked to start walking with their left legs. A total of 100 walking trials was collected.

Ethical approval for this research is granted by Curtin University ethical review committee with approval number HRE2017-0834.

IV. Data Processing

A. IMU data processing

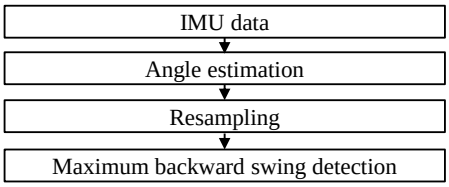
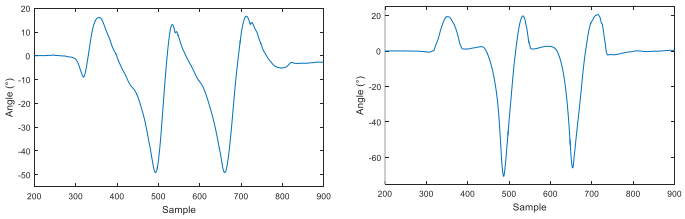


Fig. 4. IMU data processing.

Fig. 4 shows the flowchart to detect maximum backward swing of leg using IMU.



(a) (b)

Fig. 5. Estimated angle. (a) Shank angle. (b) Foot angle.

Fig. 5 shows the shank and foot angle estimated using IMU data. The angle estimation algorithm used in this research is based on our previous research [28]. This angle estimation algorithm has been validated against gold standard Vicon optical motion capture system.

As the IMU is sampled at 100Hz while the video is captured at 30Hz, we need to resample the estimated angle to 30Hz so that the IMU and video can be time-synchronized. The total number of resampled angles can be calculated according to (1). The angle can be resampled according to (2).

$$\bar{N}_{IMU} = N_{IMU} \times \frac{f_v}{f_{IMU}} = N_{IMU} \times \frac{30}{100} \quad (1)$$

$$\bar{\theta}_j = \theta_{\lceil n \rceil}(n - \lceil n \rceil + 1) - \theta_{\lfloor n \rfloor}(n - \lceil n \rceil) \quad (2)$$

$$\text{for} \begin{cases} n = 1, 1 + \frac{N_{IMU}-1}{\bar{N}_{IMU}-1}, 1 + 2\frac{N_{IMU}-1}{\bar{N}_{IMU}-1}, \dots, N_{IMU} \\ j = 1,2,3, \dots, \bar{N}_{IMU} \end{cases}$$

Where $\bar{N}_{IMU}$ denotes the total number of resampled angles, and N_{IMU} denotes the total number of samples collected by IMU. f_{IMU} and f_v denote the sampling frequency of the IMU and video camera, respectively. $\bar{\theta}$ denotes the resampled angle. $\lceil\ \rceil$ and $\lfloor\ \rfloor$ are the ceiling and flooring functions, respectively.

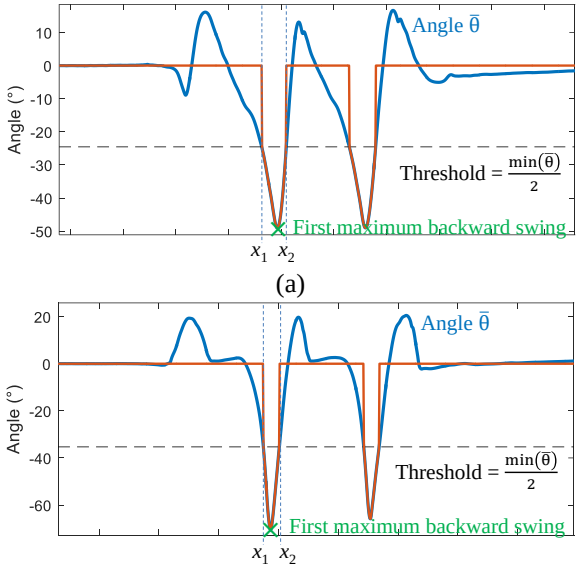


Fig. 6. First maximum backward swing detection. (a) Shank angle. (b) Foot angle.

The first maximum backward swing of the shank and foot can be detected by finding the first minimum of the shank and foot angles, respectively. Fig. 6 illustrates the first maximum backward swing can be detected by finding the minimum angle between x_1 and x_2, where x_1 and x_2 are the first and second angle that cross the threshold $\lambda_1 = \min(\bar{\theta})/2$.

B. Video Processing

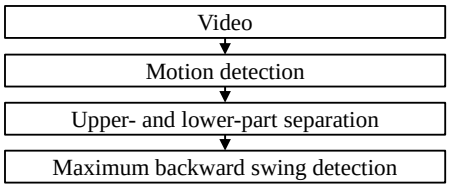


Fig. 7. Video processing flowchart.

Fig. 7 shows the flowchart to detect the maximum backward swing of leg from video.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2019.2916163, IEEE Sensors Journal

(a)

(b)

(c)

Fig. 8. Video paused at a frame. (a) RGB video v. (b) Gray-scaled video g. (c) Detected motion M in black and white.

The video captured from the smartphone is in RGB, as shown in Fig. 8(a). In order to detect the human motion, the video is first gray-scaled, as shown in Fig. 8(b), using MATLAB function 'rgb2gray' (3).

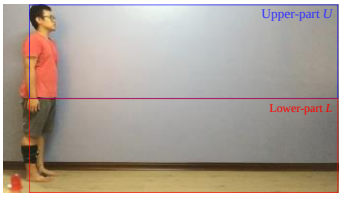
$$g_f = \text{rgb2gray}\left(v_f\right) \tag{3}$$

Where g denotes the gray-scaled video, v denotes the RGB video, and f denotes the f-th frame of the video.

The human motion can be detected by calculating the difference between the current and previous frames, and then convert it to black and white using MATLAB function 'im2bw' (4). Fig. 8(c) shows the detected motion.

$$M_f = \text{im2bw}\left(\left(g_f - g_{f-1}\right)^2, \lambda_2\right) \tag{4}$$

Where M denotes the detected motion in black and white, and λ_2 denotes the threshold. In this research, λ_2 is set to be 0.95.



(a)

(b)

Fig. 9. Upper U and lower L part separation. (a) RGB. (b) Detected motion M.

As shown in Fig. 9, the captured video is split into upper- and lower-parts equally (5) (6). This is to detect the maximum backward swing of the leg. We ignored the first 100 columns of the video because the first 100 columns are reserved for validation of proposed method using blinking LED, as stated in Section III C. In MATLAB, the first pixel was at the upper leftmost.

$$U = v_{\left(1 \text{ to } \frac{R}{2}, 100 \text{ to } C\right)} \tag{5}$$

$$L = v_{\left(\frac{R}{2} \text{ to } R, 100 \text{ to } C\right)} \tag{6}$$

Where U denotes the upper part of the video v, and L denotes the lower part of the video v. R = 720 and C = 1280 denotes the number of rows and columns of the video at any one frame, respectively.

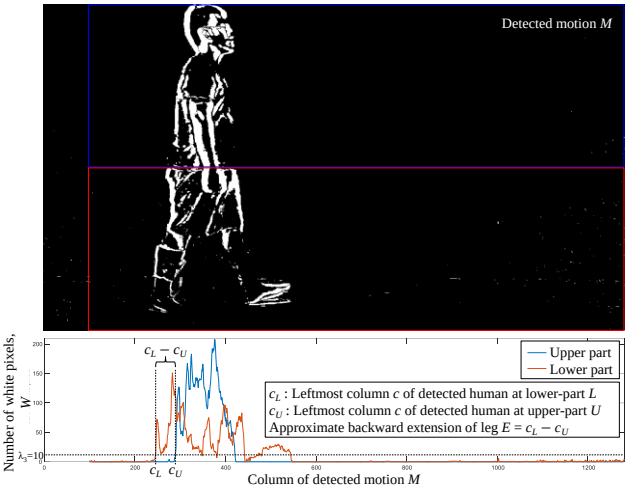


Fig. 10. Approximate extension E of leg at frame f.

Fig. 10 shows the proposed method to approximate the extension of leg in horizontal axis. The number of white pixels at each column of the detected motion M is counted by summing up M column by column (7) (8), as shown in Fig. 10.

$$W_{U,c} = \sum M_{U,c} \tag{7}$$

$$W_{L,c} = \sum M_{L,c} \tag{8}$$

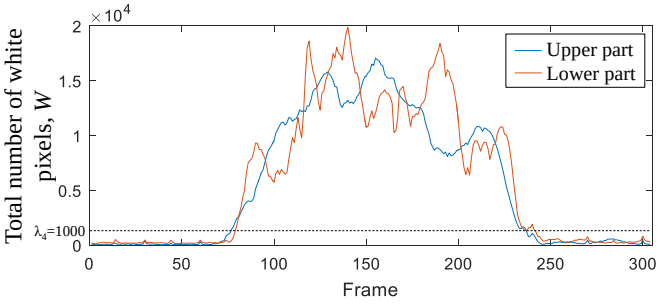
Where M_U and M_L are the detected motion M at upper U and lower L parts, respectively. $W_{U,c}$ and $W_{L,c}$ denote the total number of white pixels at each column c of M_U and M_L, respectively.

The leftmost column of the detected human can be approximated by finding the first value of W that is greater than a threshold λ_3 (9) (10). As shown in Fig. 10, λ_3 can be safely set as 20 in this research, so that the background reflections can be ignored.

for (c = 100 to 1279) **do** (9)
  if $W_{U,c} > \lambda_3$ **and** $W_{U,c+1} > \lambda_3$ **then**
    $c_U = c$
    break
  end if
end for

for (c = 100 to 1279) **do** (10)
  if $W_{L,c} > \lambda_3$ **and** $W_{L,c+1} > \lambda_3$ **then**
    $c_L = c$
    break
  end if
end for

Where c_U and c_L denote the leftmost column of detected human at upper U and lower L part, respectively.



This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2019.2916163, IEEE Sensors Journal

Fig. 11. Total number of white pixels W in the detected motion M at all frames.

In cases where the motion is very low, such as the person is standing still at frame f in Fig. 9, c_U and c_L follow the previous values at frame $f-1$. The initial c_U and c_L at the first frame are both set as 0. As shown in Fig. 11, the motion can be considered low when the total number of white pixels is lower than a threshold $\lambda_4 = 1000$.

if $\sum W_{U,f} < \lambda_4$ **then** (11)

$c_{U,f} = c_{U,f-1}$

end if

if $\sum W_{L,f} < \lambda_4$ **then** (12)

$c_{L,f} = c_{L,f-1}$

end if

Where $\sum W_{U,f}$ and $\sum W_{L,f}$ are the total number of white pixels in M_U and M_L at frame f, respectively.

The backward extension E of the leg is estimated by finding the difference between c_U and c_L (13).

$$E = c_L - c_U \quad (13)$$

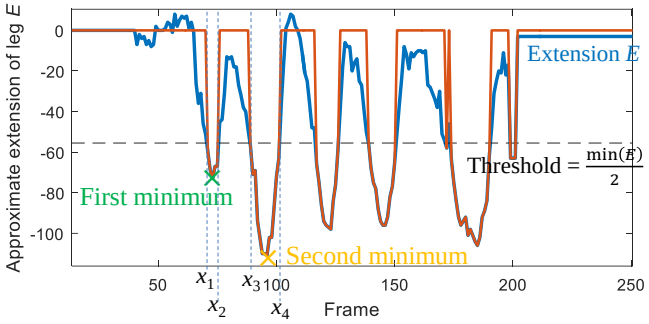


Fig. 12. Estimated maximum backward swings of leg at all frames.

Fig. 12 shows the approximated extension of leg at all frames based on proposed method. The first minimum of E can be detected by finding the minimum value between x_1 and x_2, while the second minimum of E can be detected by finding the minimum value between x_3 and x_4, where x_1 to x_4 denote the first to the forth crossing of the threshold $\lambda_5 = \min(E)/2$.

C. *Synchronization of video and IMU data*

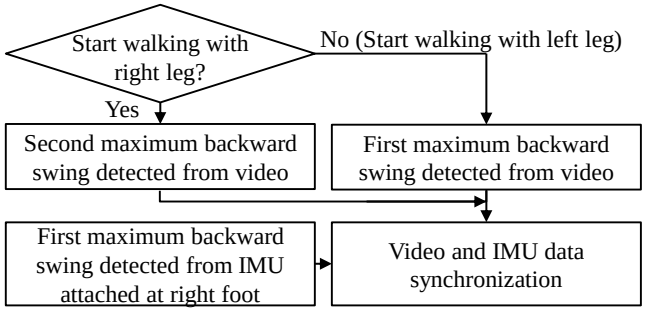


Fig. 13. Flowchart to synchronize video and IMU data.

Fig. 13 shows the flowchart to synchronize video and IMU data. As the approximated leg extension E includes both the extension of right and left legs, we need to identify whether the person starts to walk with right or left leg before synchronization.

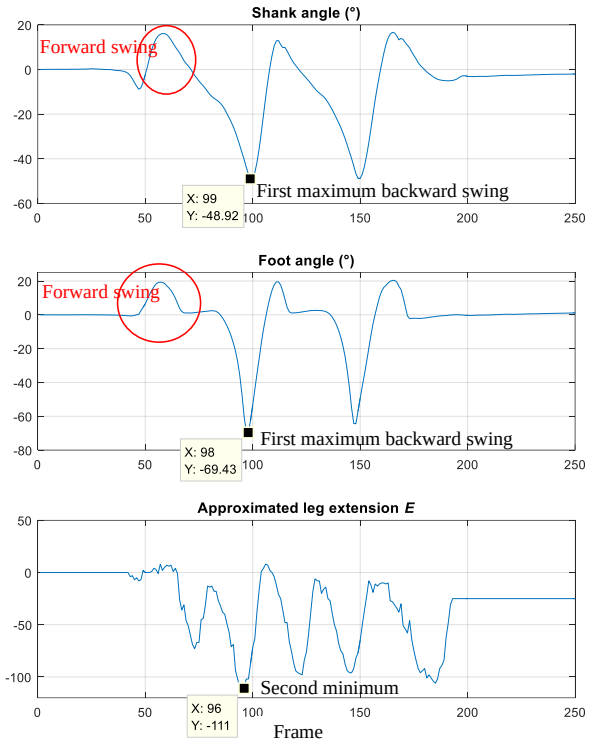


Fig. 14. Synchronize video and IMU data using the second minimum of E when the person starts to walk with right leg.

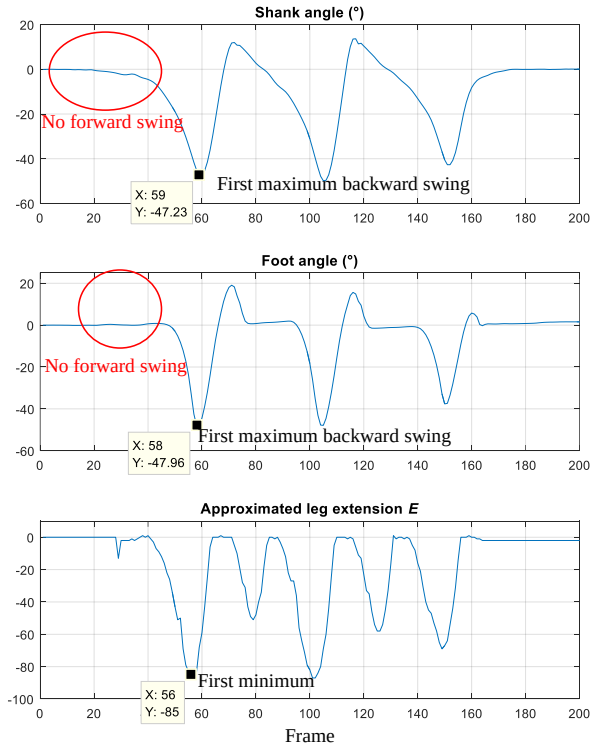


Fig. 15. Synchronize video and IMU data using the first minimum of E when the person starts to walk with left leg.

When the IMU is attached to the right leg, and the person starts to walk with right leg, there is a forward swing before the first maximum backward swing, as shown in the shank and foot

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2019.2916163, IEEE Sensors Journal

angle waveforms in Fig. 14. However, as shown in Fig. 15, when the person starts to walk with left leg, there is no forward swing before the first maximum backward swing.

Therefore, we find the maximum angle between the first sample and the maximum backward swing *B*. When there is no forward swing before the first maximum backward swing, the maximum angle between the first sample and *B* approaches zero (15).

if $\max(\bar{\theta}_{1 \text{ to } B}) \rightarrow 0$ **then** (15)

The person starts to walk with left leg.

else

The person starts to walk with right leg.

end if

The video v and IMU resampled angle $\bar{\theta}$ can be synchronized such that the video at frame A is synchronized with the IMU at sample B minus a constant k (16). The constant k is to reduce the error. We found out that using the validation method in Section IV D, k = 3 for shank angle, while k = 2 for foot angle. Fig. 14 and Fig. 15 show the synchronized video and IMU data.

$$v_{A+i} \equiv \bar{\theta}_{B-k+i} \quad \text{for } i = 0, 1, 2, \ldots \tag{16}$$

If the person starts to walk with right leg, A = second minimum of E. If the person starts to walk with left leg, A = first minimum of E. The symbol '≡' denotes synchronization.

In (2), the IMU signal is down-sampled to 30Hz. This sampling rate is considered fast enough for human's visual inspection. However, in case some applications such as automatic activity recognition may require higher sampling rate, the original angle θ can be synchronized with the video starting from sample C (17).

$$C = (B - k) \times \frac{f_{IMU}}{f_v} = (B - k) \times \frac{100}{30} \tag{17}$$

D. Validation of Proposed Method using Blinking LED

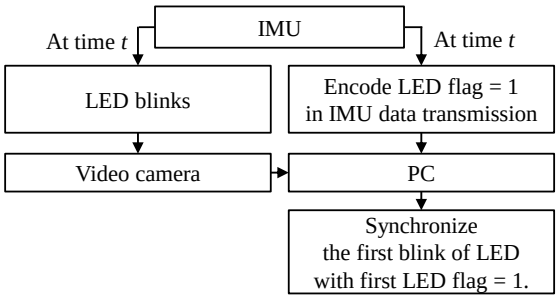


Fig. 16. Blinking LED to evaluate proposed method.

Inspired from the method in [24] which used blinking LEDs and RF transmitters to synchronize video and markers, we utilized a blinking LED and a RF transmitter to synchronize video and IMU data, for validation of our proposed method.

As stated in Fig. 16, the IMU blinks LED and at the same time, encodes LED flag = 1 in its data transmission to the PC. The video and IMU data can then be synchronized using the first blink of the LED and the first LED flag = 1.

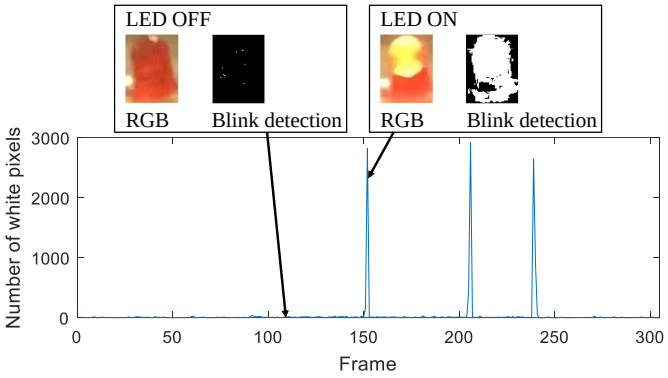


Fig. 17. LED blink detection.

The motion detection method in (4) is used to detect the blinking of LED. As shown in Fig. 17, when the LED is off, the number of white pixels in the detected motion is very low. When the LED blinks, the number of white pixels is very high. Therefore, the blinking of LED can be detected when the number of white pixels is more than a threshold λ_6 = 500.

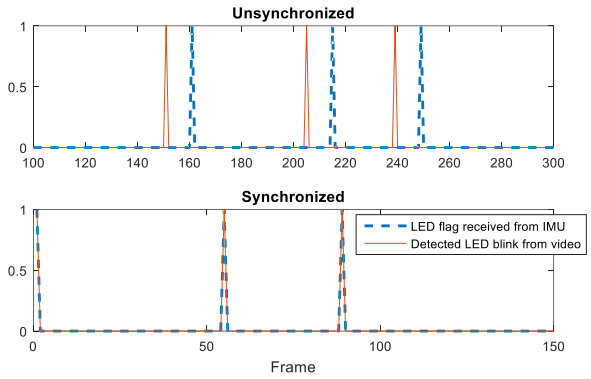


Fig. 18. Synchronize video and IMU data using first LED flag and the first blink of LED.

The video and IMU data can then be synchronized such that the first LED flag is matched with the first blink of LED, as shown in Fig. 18.

V. Results and Discussion

TABLE I
AVERAGE SYNCHRONIZATION ERROR (IN FRAMES) OF PROPOSED METHOD.

Subject	Based on Shank Angle		Based on Foot Angle	
	Mean absolute error	Max error	Mean absolute error	Max error
1	0.8	1.0	0.5	2.0
2	0.9	2.0	1.0	2.0
3	1.0	2.0	1.0	2.0
4	0.6	2.0	0.3	2.0
5	0.7	2.0	1.4	3.0
6	1.1	2.0	1.3	2.0
7	0.3	1.0	0.5	2.0
8	0.9	1.0	0.8	2.0
9	0.7	2.0	1.3	2.0
10	1.0	1.0	1.0	2.0
Average	0.8	1.6	0.9	2.1

Table I shows the synchronization error of the proposed method based on the maximum backward swings of leg detected using shank and foot angles. The synchronization error (in frames) is the difference between the video frame with the first blink of LED and the synchronized time generated by the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2019.2916163, IEEE Sensors Journal

proposed algorithm. The synchronization of video and IMU data using either shank or foot angles is very reliable with low mean absolute error of less than 1 frame. The maximum error is also low at about 2 frames.

As shown in the third plots of Fig. 14 and Fig. 15, the extension of leg *E* approximated in this research is noisy, but it can be used to detect the maximum backward swings of legs accurately.

TABLE II
COMPARISON OF VIDEO-INERTIAL SENSOR SYNCHRONIZATION METHODS AMONG LITERATURES.

Reference	Method	Mean absolute error (in frames)
[19]	Hardware	Not reported
[20]	Estimate camera-PC and sensor-PC transmission delay	<1
[21]	Cross-correlation	0.5 (horizontal gesture)
		>10 (circular gesture)
[29]	Record sensor data right after Kinect SDK software receives Kinect signal	Not reported
Proposed Method	Maximum backward swing detection	0.8 – 0.9

Table II shows several existing methods to synchronize video and inertial sensors. Bae *et al.* [19] used a synchronization hardware module to synchronize video camera and gyroscope signal. The main limitation is that the video camera and gyroscope signal must be connected through the hardware. However, in our case, the smartphone and the IMU are two separate systems and not connected together. The method by Cippitelli *et al.* [20] had achieved very low error, but it also requires an external hardware, i.e. an Arduino board connected to PC to control seven LEDs for transmission time delay estimation.

Compared to the existing methods above, an advantage of our proposed method is that it does not require external device or LED for synchronization. Validated against the LED-blinking method, our proposed method achieves very low synchronization error at an average of 0.9 frames. There are also existing synchronization methods which do not require external hardware. Plotz *et al.* [21] used cross-correlation method to synchronize video and accelerometer data. Although horizontal hand gesture had very low synchronization error, the circular hand gesture had very high error. Liu *et al.* [29] recorded the inertial sensor data right after the Kinect SDK software received signal from the Kinect camera. Although [29] did not report the synchronization error, the method most likely consists of synchronization error due to the transmission time delay as stated in [20].

TABLE III
AVERAGE EXECUTION TIME AND SYNCHRONIZATION ERROR OF PROPOSED METHOD AT DIFFERENT FRAME RATES.

Frame rate (frame/s)		15	30
Average execution time (s)		1.12	2.05
Mean absolute synchronization error (ms)	Shank	32	27
	Foot	25	30

The original video (30 frame/s) is down-sampled to 15 frame/s for comparison.

As shown in Table III, the average execution time to synchronize the video and inertial sensor data is 1.12 and 2.05 seconds for 15 frame/s and 30 frame/s videos, respectively, while the synchronization error is almost the same. The computation time is considered long, and it is mainly because our proposed method requires all video frames to be processed to obtain the threshold = min(*E*)/2 as shown in Fig. 12. The frame-by-frame reading of the video in MATLAB consumes most of the computation time, compared to the processing of the algorithms.

There are several methods to reduce the computation time drastically, such as by calculating the threshold = min($E_{1\rightarrow N/2}$)/2 instead of threshold = min($E_{1\rightarrow N}$)/2 where *N* is the total number of frames. However, the 2-second computation time is not a problem as our method is designed to assist clinicians in walking assessment, which does not require real-time processing.

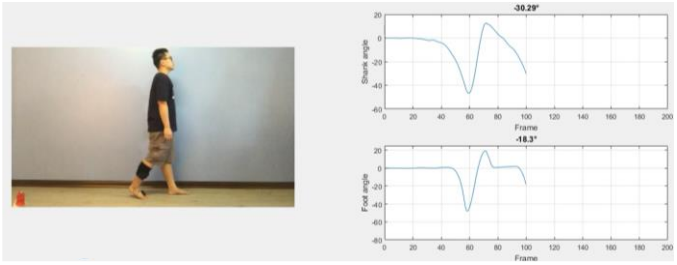


Fig. 19. Synchronized output for visual inspection paused at 100th frame.

Fig. 19 shows a paused video of a person walking with the synchronized shank and foot angle shown at the right side. This allows the clinicians to know the shank and foot angles for possibly better visual inspection of the gait. The full video of Fig. 19 is attached as multimedia with this journal.

In this research, we used only two IMUs, but more inertial sensors can be added for full body tracking as long as they are temporally synchronized. Besides, synchronized force sensitive resistors (FSRs) used in [30] can be added for gait phase detection. With 3-D angle estimation algorithm such as [14], [15], and [28], the 3-D joint angle information could also be provided to the clinicians for visual inspection. Other than providing the joint angle information, we could add gait phase information to the video [30].

A limitation of our proposed method is that the video in our proposed method must be captured from the side of the user for maximum backward swing detection. The proposed method cannot be used to detect maximum backward swing of the subject if the video is captured from the front. Besides, our method can only support only one user. If there are two people moving in the video, the proposed method cannot differentiate which user is wearing the IMUs.

VI. CONCLUSION

A method to temporally synchronize video and IMU data was proposed. The proposed method is based on motion detection, and it has achieved very low errors without accessing to the camera and inertial sensor's internal system clock. The main idea of the proposed method is to detect and match the maximum backward swing of the leg for synchronization. The synchronized joint angle information obtained from the inertial sensors can be placed side-by-side to the video for clinicians to perform better visual inspection. The main limitation is that the

video must be captured from the side of human for maximum backward swing detection. The proposed method cannot be used to detect maximum backward swing of the subject if the video is captured from the front. In future, we plan to temporally synchronize the video and IMU data when the video is captured from the front. More synchronized sensors can also be added for full body tracking.

REFERENCES

[1] Hashimoto, K., Higuchi, K., Nakayama, Y., & Abo, M. (2007). Ability for basic movement as an early predictor of functioning related to activities of daily living in stroke patients. *Neurorehabilitation and Neural Repair*, *21*(4), 353-357.

[2] A. Muro-de-la-Herran, B. Garcia-Zapirain and A. Mendez-Zorrilla. Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications. *Sensors 14(2)*, pp. 3362-3394. 2014. DOI: http://dx.doi.org.dbgw.lis.curtin.edu.au/10.3390/s140203362.

[3] T. Krosshaug *et al*, "Estimating 3D joint kinematics from video sequences of running and cutting maneuvers--assessing the accuracy of simple visual inspection," *Gait Posture*, vol. 26, *(3)*, pp. 378-385, 2007.

[4] "Motion Capture Systems", *VICON*, 2018. [Online]. Available: https://www.vicon.com/. [Accessed: 16 Dec. 2018].

[5] "OptiTrack", *OptiTrack*, 2018. [Online]. Available: https://optitrack.com/. [Accessed: 16 Dec. 2018].

[6] W. Geiger *et al*., "MEMS IMU for AHRS applications," *2008 IEEE/ION Position, Location and Navigation Symposium*, Monterey, CA, 2008, pp. 225-231.
doi: 10.1109/PLANS.2008.4569973

[7] Y. Tao, H. Hu and H. Zhou, "Integration of Vision and Inertial Sensors for 3D Arm Motion Tracking in Home-based Rehabilitation," *Int. J. Robotics Res.*, vol. 26, *(6)*, pp. 607, 2007.

[8] Teixeira T, Jung D, Savvides A (2010) Tasking networked cctv cameras and mobile phones to identify and localize multiple people. In: Proceedings of the 12th ACM international conference on Ubiquitous computing, Ubicomp '10, ACM, New York, NY, USA, pp 213–222

[9] C. Jia and B. L. Evans, "Online Camera-Gyroscope Autocalibration for Cell Phones," in *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5070-5081, Dec. 2014.
doi: 10.1109/TIP.2014.2360120

[10] E. Bertino and E. Ferrari, "Temporal synchronization models for multimedia data," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 10, no. 4, pp. 612-631, July-Aug. 1998.
doi: 10.1109/69.706060

[11] M. Meribout, M. Nakanishi and T. Ogura, "A parallel algorithm for real-time object recognition," *Pattern Recognition Journal*, vol. 35, *(9)*, pp. 1917-1931, 2002.

[12] S. Hu and H. Zhang, "Image Edge Detection Based on FCM and Improved Canny Operator in NSST Domain," *2018 14th IEEE International Conference on Signal Processing (ICSP)*, Beijing, China, 2018, pp. 363-368.
doi: 10.1109/ICSP.2018.8652426

[13] X. Zhang, M. Ding and G. Fan, "Video-Based Human Walking Estimation Using Joint Gait and Pose Manifolds," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 7, pp. 1540-1554, July 2017.
doi: 10.1109/TCSVT.2016.2527218

[14] S. O. H. Madgwick, A. J. L. Harrison and R. Vaidyanathan, "Estimation of IMU and MARG orientation using a gradient descent algorithm," *2011 IEEE International Conference on Rehabilitation Robotics*, Zurich, 2011, pp. 1-7.
doi: 10.1109/ICORR.2011.5975346

[15] Y. Zhang, K. Chen, J. Yi, T. Liu and Q. Pan, "Whole-Body Pose Estimation in Human Bicycle Riding Using a Small Set of Wearable Sensors," in *IEEE/ASME Transactions on Mechatronics*, vol. 21, no. 1, pp. 163-174, Feb. 2016.
doi: 10.1109/TMECH.2015.2490118

[16] C. Chen, R. Jafari and N. Kehtarnavaz, "UTD-MHAD: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor," *2015 IEEE International Conference on Image Processing (ICIP)*, Quebec City, QC, 2015, pp. 168-172.
doi: 10.1109/ICIP.2015.7350781

[17] A. Farnoosh, M. Nabian, P. Closas and S. Ostadabbas, "First-person indoor navigation via vision-inertial data fusion," *2018 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, Monterey, CA, 2018, pp. 1213-1222.
doi: 10.1109/PLANS.2018.8373507

[18] P. Jatesiktat, D. Anopas and W. T. Ang, "Personalized Markerless Upper-Body Tracking with a Depth Camera and Wrist-Worn Inertial Measurement Units," *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Honolulu, HI, 2018, pp. 1-6.
doi: 10.1109/EMBC.2018.8513068

[19] Jum-Han Bae and Jong-Tae Kim, "Design and implementation of high accurate synchronization between gyroscope and image sensor," *2015 Seventh International Conference on Ubiquitous and Future Networks*, Sapporo, 2015, pp. 956-958.
doi: 10.1109/ICUFN.2015.7182687

[20] T. Plotz, C. Chen, N. Y. Hammerla and G. D. Abowd, "Automatic Synchronization of Wearable Sensors and Video-Cameras for Ground Truth Annotation -- A Practical Approach," *2012 16th International Symposium on Wearable Computers*, Newcastle, 2012, pp. 100-103.
doi: 10.1109/ISWC.2012.15

[21] E. Cippitelli *et al*., "Time synchronization and data fusion for RGB-Depth cameras and inertial sensors in AAL applications," *2015 IEEE International Conference on Communication Workshop (ICCW)*, London, 2015, pp. 265-270.
doi: 10.1109/ICCW.2015.7247189

[22] X. Lin, V. Kitanovski, Q. Zhang and E. Izquierdo, "Enhanced multi-view dancing videos synchronisation," *2012 13th International Workshop on Image Analysis for Multimedia Interactive Services*, Dublin, 2012, pp. 1-4.
doi: 10.1109/WIAMIS.2012.6226773

[23] N. Q. K. Duong and F. Thudor, "Movie synchronization by audio landmark matching," *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, BC, 2013, pp. 3632-3636.
doi: 10.1109/ICASSP.2013.6638335

[24] Y. K. Ryu and C. Oh, "RF Signal Synchronized Low Cost Motion Capture Device," *2007 Asia-Pacific Microwave Conference*, Bangkok, 2007, pp. 1-4.
doi: 10.1109/APMC.2007.4554680

[25] F. Ofli, R. Chaudhry, G. Kurillo, R. Vidal and R. Bajcsy, "Berkeley MHAD: A comprehensive Multimodal Human Action Database," *2013 IEEE Workshop on Applications of Computer Vision (WACV)*, Tampa, FL, 2013, pp. 53-60.
doi: 10.1109/WACV.2013.6474999

[26] *Invensense*, 2019. [Online]. Available: www.invensense.com. [Accessed: 4 Jan. 2019].

[27] "Compare iPhone Models", *Apple*, 2019. [Online]. Available: https://www.apple.com/my/iphone/compare/. [Accessed: 16 Jan. 2019].

[28] Y. C. Han, K. I. Wong and I. Murray, "2-Point Error Estimation Algorithm for 3-D Thigh and Shank Angles Estimation Using IMU," in *IEEE Sensors Journal*, vol. 18, no. 20, pp. 8525-8531, 15 Oct.15, 2018.
doi: 10.1109/JSEN.2018.2865764

[29] K. Liu, C. Chen, R. Jafari and N. Kehtarnavaz, "Fusion of Inertial and Depth Sensor Data for Robust Hand Gesture Recognition," in *IEEE Sensors Journal*, vol. 14, no. 6, pp. 1898-1903, June 2014.
doi: 10.1109/JSEN.2014.2306094

[30] Y. C. Han, K. I. Wong and I. Murray, "Gait Phase Detection for Normal and Abnormal Gaits Using IMU," in *IEEE Sensors Journal*, vol. 19, no. 9, pp. 3439-3448, 1 May1, 2019.
doi: 10.1109/JSEN.2019.2894143