Robotica (2019) volume 37, pp. 928–946.
doi:10.1017/S0263574717000571

A comparative study of in-field motion capture approaches for body kinematics measurement in construction

JoonOh Seo†,*, Abdullatif Alwasel‡, SangHyun Lee§, Eihab M. Abdel-Rahman‡ and Carl Haas¶

†*Department of Building and Real Estate, Hong Kong Polytechnic University, Hong Kong*
‡*Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada*
§*Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI, USA*
¶*Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, ON, Canada*

(Accepted November 11, 2017. First published online: December 20, 2017)

SUMMARY
Due to physically demanding tasks in construction, workers are exposed to significant safety and health risks. Measuring and evaluating body kinematics while performing tasks helps to identify the fundamental causes of excessive physical demands, enabling practitioners to implement appropriate interventions to reduce them. Recently, non-invasive or minimally invasive motion capture approaches such as vision-based motion capture systems and angular measurement sensors have emerged, which can be used for in-field kinematics measurements, minimally interfering with on-going work. Given that these approaches have pros and cons for kinematic measurement due to adopted sensors and algorithms, an in-depth understanding of the performance of each approach will support better decisions for their adoption in construction. With this background, the authors evaluate the performance of vision-based (RGB-D sensor-, stereovision camera-, and multiple camera-based) and an angular measurement sensor-based (i.e., an optical encoder) approach to measure body angles through experimental testing. Specifically, measured body angles from these approaches were compared with the ones obtained from a marker-based motion capture system that has less than 0.1 mm of errors. The results showed that vision-based approaches have about 5–10 degrees of error in body angles, while an angular measurement sensor-based approach measured body angles with about 3 degrees of error during diverse tasks. The results indicate that, in general, these approaches can be applicable for diverse ergonomic methods to identify potential safety and health risks, such as rough postural assessment, time and motion study or trajectory analysis where some errors in motion data would not significantly sacrifice their reliability. Combined with relatively accurate angular measurement sensors, vision-based motion capture approaches also have great potential to enable us to perform in-depth physical demand analysis such as biomechanical analysis that requires full-body motion data, even though further improvement of accuracy is necessary. Additionally, understanding of body kinematics of workers would enable ergonomic mechanical design for automated machines and assistive robots that helps to reduce physical demands while supporting workers' capabilities.

KEYWORDS: Body kinematics; Motion capture; Construction

Introduction

Construction workers are frequently exposed to excessive physical demands in a relatively dangerous and unhealthy working environment that may lead to health and safety issues.[1] Despite recent improvements by adopting best practices (e.g., tool box meetings and structured hazard analysis

* Corresponding author. E-mail: joonoh.seo@polyu.edu.hk

processes) to deal with these issues, the construction sector still remains as one of the hazardous industries, showing higher rates of fatal and nonfatal injuries than other industries.[2] Consequently, evaluating and controlling physical demands from the job, equipment and environment not to exceed one's capabilities is essential to prevent and mitigate potential health and safety risks.

In-depth evaluation and control of physical demands should begin with measuring body kinematics that include body position, displacement, velocity and acceleration.[3] Human motions not only create loads on the involved musculoskeletal system (e.g., muscles and tendons) by themselves or combined with external forces, but they also have an important role in accompanying an action by transmitting forces generated from a body (i.e., active muscle contraction) to the external environment.[4] As a result, kinematics data can provide contextual information on the fundamental causes of changing physical demands as a worker's behavior is affected by physical work environmental factors (e.g., geometry of the workplaces, temperature, and types of tools), as well as individual factors (e.g., anthropometry and preferred working techniques).[5] Also, enhanced awareness of body kinematics of workers helps to support effective design of automated machines and assistive robotics that can not only reduce physical demands from work, but also enhance workers' capabilities by improving ergonomics in human-machine (or robot) interaction.[6,7] Such an understanding of workers body kinematics would also inform contemporary research in construction ergonomics and robotics, given current developments in architecture, gerontechnology, exoskeletons, anthropomorphic robotics, augmented reality, and industrialized construction environments because they all demand highly interdependent and integrated kinematics systems solutions.[9–15]

Generally, measuring body kinematics is enabled by using motion capture techniques that are based on optical, inertial, mechanical, magnetic or acoustic approaches. Among them, optical motion capture systems are the most popular solution to obtain three-dimensional (3D) full-body motion data by tracking active or passive markers attached to the body. They have been widely used for diverse applications including clinical motion analysis and biomechanical study. Despite their precision and reliability, their applications in practice have been limited due to the need for complex laboratory settings, the high cost of devices and the time-consuming procedure for data collection. Their most critical drawback is the need for simulating tasks in a controlled environment by assuming that captured motions correspond to typical activities under real conditions.[8] However, considering the continuously changing and unstructured nature of working environments in construction, it is hard to simulate construction tasks in a controlled setting by reflecting all possible situations that would exist on sites. As a result, an effective and easily accessible means for collecting in-field motion data at construction sites is required to evaluate potential safety and health risks of workers while performing tasks.

Recently, vision-based (i.e., markerless) motion capture approaches have gained interest. These approaches appear to be promising as an in-field motion data collection method by addressing some of the major challenges that exist with optical motion capture systems. For example, as these vision-based approaches obtain body kinematics data by processing two-dimensional (2D) or 3D images directly collected from real conditions, they do not need complex laboratory settings or markers attached to a human body. Also, 2D or 3D images can be collected using any type of existing ordinary video cameras or affordable 3D image-sensing devices (e.g., RGB-D (Red, Green, Blue and Depth) sensors and stereovision cameras). The ease of use, non-invasiveness and cost-effectiveness of vision-based approaches can broaden the spectrum of their applications for job analysis under real conditions. Additionally, the use of body fixed sensors such as Inertial Measurement Units (IMUs) or angular measurement sensors (e.g., goniometers, optical encoders, strain gauges or magnetic sensors) has provided effective solutions for in-field motion measurement.[8] Combined with a wireless data transmission capability, these approaches allow us to obtain real-time motion data. For 3D full-body kinematic measurements, several commercialized IMU-based (e.g., Xsens™ (Xsens North America, Inc., xsens.com)) or mechanical (e.g., Gypsy 7™ (Meta Motion, metamotion.com)) motion capture systems are available, but the need for wearing a full-body suit equipped with sensors may interfere with on-going work. Instead, wearing light-weight and wearable angular measurement sensors attached only at a specific body joint of interest may minimize workers' discomfort during performing tasks, though they only provide one degree of freedom joint angle.

To measure workers' body kinematics non-invasively or minimally invasively at construction sites, both vision-based and angular measurement sensor-based approaches are viable means, although several limitations such as sensitivity to self-occlusions (i.e., occlusions of specific body joints

https://doi.org/10.1017/S0263574717000571 Published online by Cambridge University Press

by other body parts) of vision-based approaches and limited use of angular measurement sensors still remain.[8,16–18] Given the pros and cons of these approaches, an in-depth understanding of the performance of each approach can lead to better decisions for appropriate uses of these approaches in construction.

With this background, this paper reports on the evaluation of motion data obtained from vision-based and angular measurement sensor-based approaches through an experimental study. Specifically, three emerging vision-based approaches for collecting motion data during construction tasks are selected. Those are: (1) RGB-D sensor-based;[17] (2) stereo-vision camera-based;[19] and (3) multiple camera-based[16,18] approaches. An optical encoder, which is a potentiometer-based electro-goniometers,[20] is applied as an angular measurement sensor. Also, a marker-based motion capture system (Optotrak™, Northern Digital, Inc., Waterloo, Canada) is used as the ground truth of motion data. To compare the accuracy of motion data, selected joint angles from each approach are compared with the ones from Optotrak™ during performance of several dynamic tasks. Based on the results, the performance of these approaches and their potential application areas for analyzing construction tasks are discussed.

State-of-the-Art in In-Field Body Kinematics Measurements

This section describes the technical aspects and procedures of the state-of-the-art approaches that enable us to non-invasively or minimally invasively measure body kinematics while workers perform tasks at construction sites. Those include: (1) vision-based approaches, and (2) angular measurement sensor-based approaches. Also, by reviewing previous work on these approaches, the pros and cons of each approach are summarized.

Vision-based motion capture approaches

Vision-based approaches aim to extract full-body motion data by processing 2D or 3D images.[21] Previous research efforts have developed several vision-based approaches: (1) RGB-D sensor-based;[17,22–25] (2) stereo-vision camera-based;[19,53] and (3) multiple camera-based[16,18] approaches. While RGB-D sensor- and stereovision camera-based approaches take an advantage of the 3D imaging sensors that directly provide 3D information on scenes, a multiple camera-based approach relies on photogeometric acquisition of 3D body joint locations (i.e., 3D reconstruction) from tracked 2D joint locations of multi-view images.

RGB-D sensor-based approach. Several computer vision algorithms have been developed to estimate human poses by detecting the 3D positions of body joints directly from RGB-D images.[22–25] Recently, motion capture solutions such as iPi Desktop Motion Capture (www.ipisoft.com) and OpenNI (www.openni.org) that use a Microsoft Kinect sensor have provided effective solutions for extracting skeleton-based motion data from 3D images obtained by RGB-D sensors.

The Kinect sensor that was initially developed for video gaming is capable of providing both depth and color information at the resolution of 640×480 and the rate of 30 frames per second (fps).[26] This sensor is equipped with an infrared (IR) projector, a color camera and an IR camera. Using the projected structured IR lights, it measures the depth, reconstructing 3D scenes with point cloud.[27] Combined with the 3D sensing feature of the Kinect sensor, the iPi Desktop Motion Capture software provides a marker-less solution for collecting full-body motion data. Figure 1 shows an example of an RGB-D image with a pre-defined body model, and the corresponding motion data. Basically, the algorithm in this software is model based, which means that motion data can be tracked by matching the surface of a pre-defined body model with a depth image (Fig. 1A). Then, the tracked motion data can be exported into any types of motion data formats such as the Biovision Hierarchy (BVH) motion data (Fig. 1B). This software also provides several post-processing algorithms to refine tracking and filtering algorithms for noise removal and smoothing.

The following are the advantages of an RGB-D sensor-based motion capture approach: (1) no need for markers or sensors attached to human body, which allows for motion capture without interfering with on-going work; (2) low cost (e.g., approximately 150–250 USD); (3) an easy-to-use and easy-to-carry means for in-field motion data collection.[17] Technically, this approach is robust to self-occlusions because the iPi software provides an inverse kinematics algorithm that can adjust incorrectly tracked body parts due to occlusions. However, as the Kinect sensor uses IR light, the use

https://doi.org/10.1017/S0263574717000571 Published online by Cambridge University Press

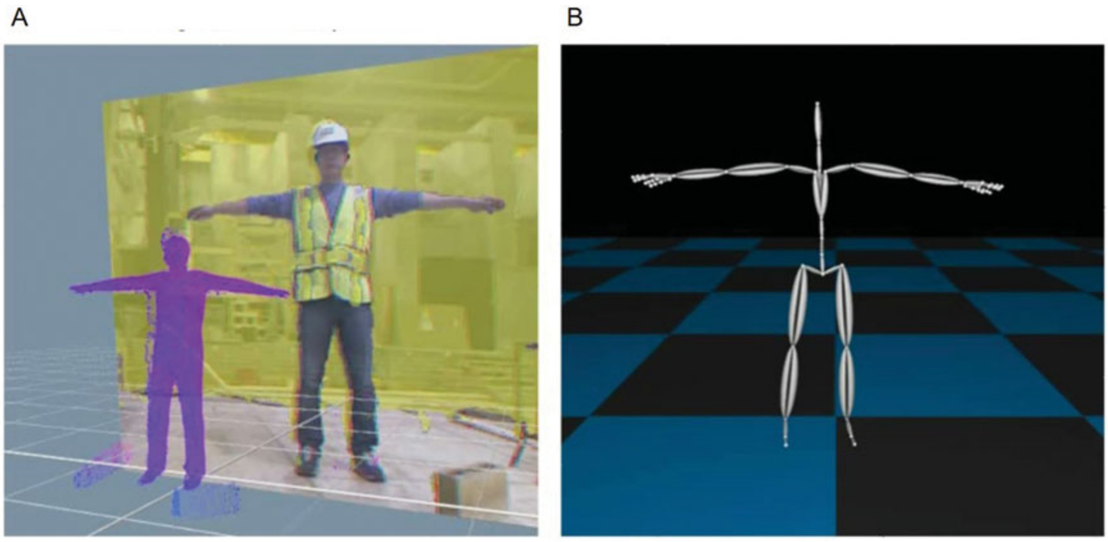

Fig. 1. RGB-D sensor (i.e., Kinect™)-based motion capture. (A) RGB-D image with a body model. (B) Skeleton-based motion data (BHV).

of this approach is limited only in an indoor environment due to its sensitivity to sunlight. Also, the short operating range of the Kinect sensor (within 4 m) is one of the disadvantages of this approach.

Stereovision camera-based approach. A stereovision system is designed to extract 3D information from a stereo image pair.[28] Stereovision works in a similar way to 3D sensing in human vision. It begins with identifying image pixels that correspond to the same point in a physical scene observed by multiple cameras. The 3D position of a point can then be established by triangulation using a ray from each camera. The more the corresponding pixels identified, the more the 3D points that can be determined with a single set of images. Correlation stereo methods attempt to obtain correspondences for every pixel in the stereo image, resulting in tens of thousands of 3D values generated with every stereo image. Bumblebee XB3™ manufactured by Point Grey Technologies (www.ptgrey.com) is one of the widely used stereovision cameras. The stereo camera measures line-of-sight distance using two lenses with a narrow baseline in a self-contained unit. This allows for both optical and depth data to be collected with a few environmental restrictions (e.g. outdoor environments) and limited field-of-view.

Starbuck *et al.*[19] proposed a stereovision camera-based motion capture approach that addresses the short operating range of an RGB-D sensor. The 3D point cloud data collected from the stereovision camera was converted into a format used by an existing kinematic modeling software solution (i.e., iPi Motion Capture software) designed for use with RGB-D sensors. Then, using the same algorithm used in an RGB-D sensor-based approach, skeleton-based motion data was extracted from the 3D point cloud data. Through a laboratory test, the proposed method was proved to be comparable to the traditional RGB-D sensor-based approach.[19]

A stereovision camera-based approach provides additional advantages, beyond the benefits from the RGB-D sensor-based approach. For example, the use of this approach does not suffer from environmental conditions, allowing both indoor and outdoor applications. Also, the operating range of the stereovision camera is flexible according to lens field-of-view, lens separation, and image size.[29] However, as computing depth information from two images is a computationally intensive task, the frame rate relies on the performance of hardware.[29]

Multiple camera-based approach. A multiple camera-based motion capture approach aims to estimate the 3D locations of body joints by processing 2D images from two different views using multiple video cameras or a 3D camcorder that has two lenses in one camera.[16] Han and Lee[16] proposed a motion capture process that consists of: (1) 2D pose estimation from one view of images; (2) correspondence matching of body joints on the other view of images; and (3) 3D reconstruction of body joints using the corresponding joint locations identified. However, this approach suffered

https://doi.org/10.1017/S0263574717000571 Published online by Cambridge University Press

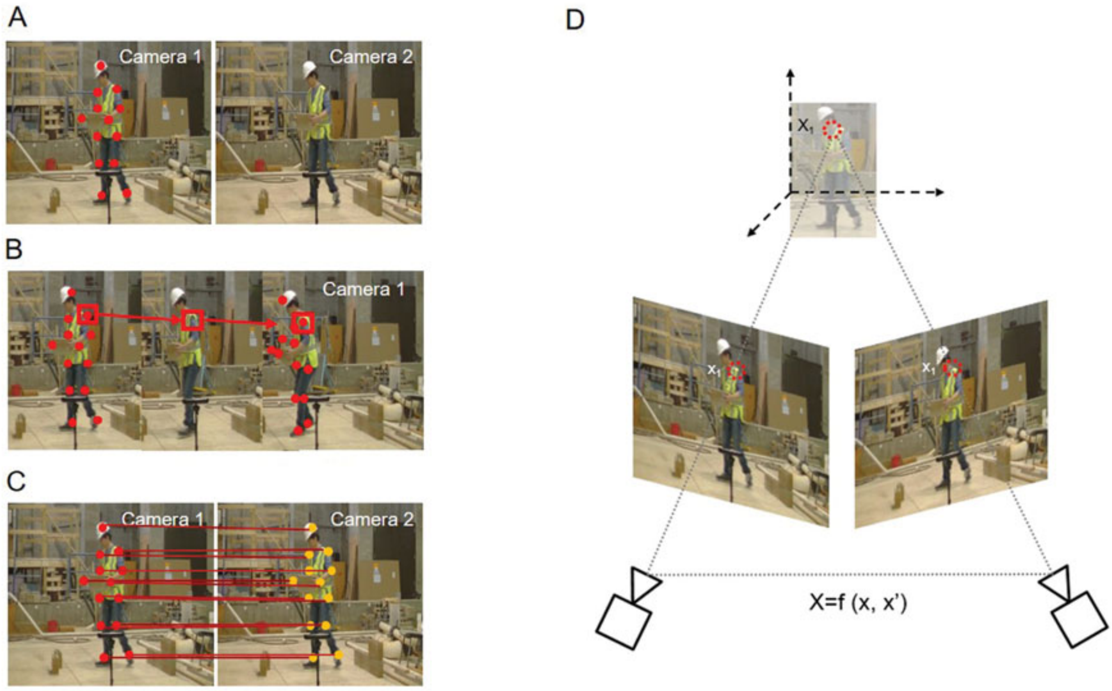

Fig. 2. An overview of a multiple camera-based motion capture approach. (A) Initialization. (B) Body joint tracking. (C) Correspondence matching. (D) 3D reconstruction.

from the need for extensive training images to detect joint locations on testing images, and significant computation time for 2D pose estimation. To address this issue, Liu *et al.*[18] modified Han and Lee's[16] approach by proposing body joint tracking that accelerates the 2D pose estimation process without the prior knowledge (training images for joint detection). Figure 2 shows an overview of the modified approach.

The main idea of 2D joint tracking is that continuous tracking of body joints on consecutive image frames enables fast estimation of 2D skeletons.[18] Once the target joints are initialized in the first frame (Fig. 2A), the algorithm tracks the joints in consecutive images by detecting the image patch with the most similar color histogram with that of the initialized target. To reduce computation time, a modified particle filter tracker was applied to specify a number of reliable candidates for the targets in the subsequent frames.[30] The tracking of different body joints is performed independently, resulting in a 2D skeleton model (Fig. 2B). The next process is to identify the corresponding body joints on the image from the other viewpoint by comparing the features of a pixel with the feature descriptors such as SIFT (Scale-Invariant Feature Transform)[31] and SURF (Speeded Up Robust Features)[32,33] (Fig. 2C). To obtain more reliable corresponding locations of body joints, the search space is constrained by epipolar geometry[34] and homography.[16] Once pairs of corresponding body joints are detected from two different viewpoints of images, a 3D reconstruction algorithm detects the 3D positions of each joint through triangulation, resulting in 3D full-body skeleton-based motion data as shown in Fig. 2D. Camera intrinsic and extrinsic parameters required for 3D reconstruction are obtained by using Zhang's[35] camera calibration technique.

The strength of a multiple camera-based motion capture approach is that we can use ordinary video cameras to obtain motion data, and thus this approach is less hardware dependent than RGB-D sensor- and stereovision camera-based approaches. Also, it is not only cost effective, but also we can benefit from zoom lenses that collect video images from a distance. Even though the environmental conditions such as illuminations may affect the performance of 3D skeleton extraction, post-image processing enables us to obtain clear images even in a noisy environment. From previous studies that investigated the accuracy of this approach, about $\pm$ 10 cm of errors in body length and up to 20 degrees of errors in joint rotation angles have been reported.[16,18] These errors came from either incorrectly detected joint locations or an inaccurate camera calibration process. Especially, the performance of this approach was significantly affected by frequent self-occlusions of forearms (e.g., elbows and hands), which led to larger errors.[16,18]

Angular measurement sensor-based approaches

Angular measurement sensor-based approaches directly measure joint angles using sensors attached to specific body joints without the need for any mathematical transformation in space or time. Examples of sensors include goniometers and strain gauges.

Goniometers. Goniometers have been used to measure joint range of motion. Traditional goniometers were made of a mechanical compass that measures the static relative angle between two body segments.[36] Modern goniometers are made of an electrical compass (potentiometer-based) that can measure static and dynamic relative angles.[36,37] The potentiometer changes its resistance with the rotation of the two body segments connected to it. The principle of operation is that the voltage drop (V) across the potentiometer due to a constant electric current (I) passing through it will depend on the resistance (R) following Ohm's law:

$$V = IR \tag{1}$$

Calibration of the potentiometer (goniometer) from 0° to a full range of motion is conducted once to produce a calibration chart that describes the relationship between the change in joint angle and the measured voltage. The use of potentiometer allows for detection of rotary motion as well as for placement of the sensor at center of joint rotation.

Strain gauges. Strain gauges work on the same principles as goniometers except that the sensing element in a strain gauge responds to translation (change in length (ΔL)) represented as change in resistance (ΔR):

$$\Delta R/R = \Delta L/L \tag{2}$$

Measuring the joint angle depends on the placement of the strain gauge with respect to the axis of joint rotation.[38] Misalignment can produce significant errors due to the complexity of placing a translational sensor to detect rotatory motion.[39]

Experimental Comparison of In-Field Motion Capture Approaches

The section describes an experimental test to compare the accuracy of three vision-based motion capture approaches and one angular measurement sensor-based approach: (1) an RGB-D sensor-based approach; (2) a stereovision camera-based approach; (3) a multiple camera-based approach; and (4) an optical encoder (i.e., a potentiometer-based electrogoniometer).

Testing conditions

Vision-based motion capture approaches and an angular measurement sensor-based approach were tested through two independent testing sessions as shown in Fig. 3. An exoskeleton is used to align the optical encoder with the knee flexion axis of rotation. The straps used to attach the exoskeleton to the lower limb were indistinguishable from the subject's clothes and skin. As a result, they could affect performance of image processing for the vision-based approaches, especially the multiple camera-based approach that tracks body joints using color information. To avoid this, the angular measurement sensor-based approach was tested in a separate session from the vision-based approaches.

Figure 3 shows the experimental conditions for each testing session. In the session for vision-based approaches (Fig. 3A), three image sensors were located in front of a subject to collect 2D or 3D images from a front view. The Kinect™ sensor (640×480 resolution with 30 fps), Bumblebee XB3™ stereovision camera (320×240 resolution with about 10 fps) and 3D camcorder (1920×1080 resolution with 29 fps) were positioned 4, 6 and 8 m away from the subject, respectively. The positions of Kinect™ and Bumblebee XB3™ were determined based on the optimal operating distance proposed by the manufacturer. As the 3D camcorder has zoom lenses, its position was selected to obtain a clear view of the subject's whole body. Motion data obtained from an Optotrak™ system served as a ground truth. Optotrak™ uses active markers attached on the center of body joints to track body motions. If the markers are captured by at least one of cameras, the system can provide

https://doi.org/10.1017/S0263574717000571 Published online by Cambridge University Press

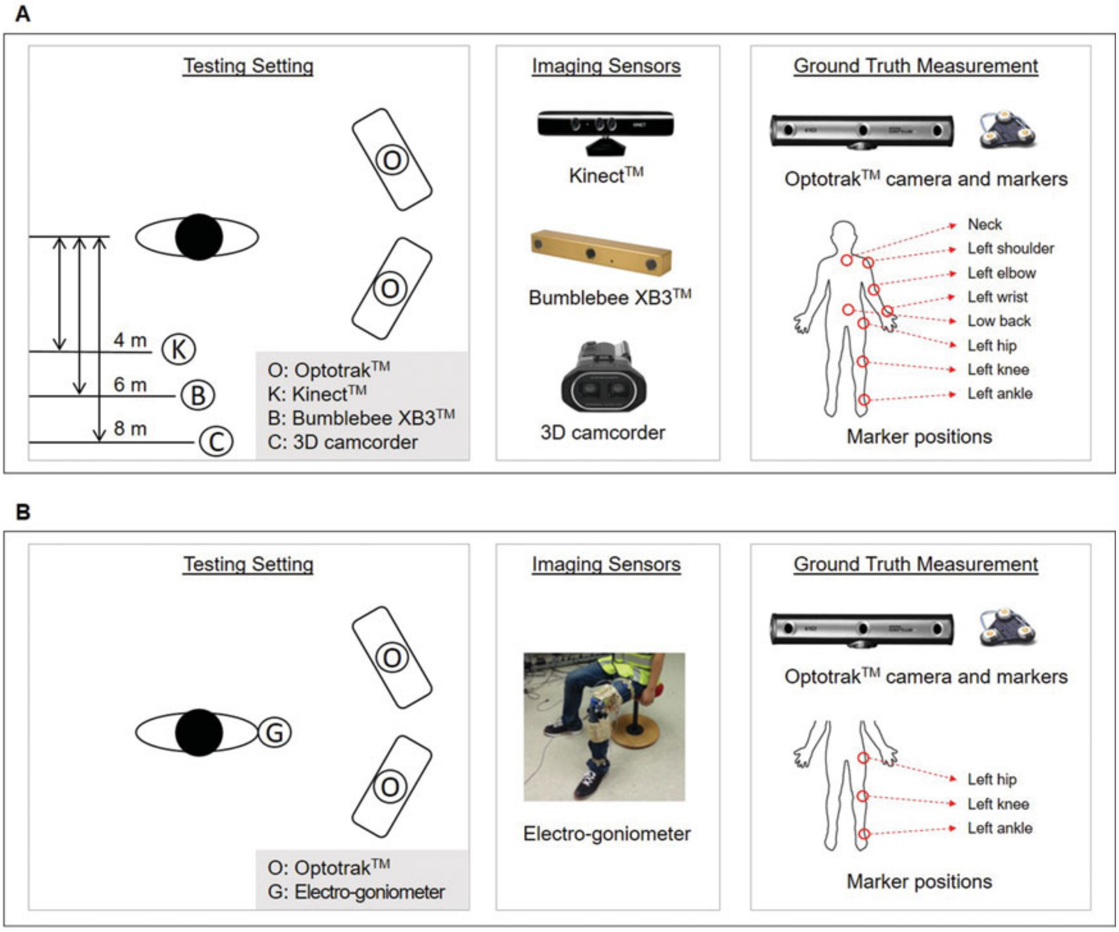

Fig. 3. Experimental settings and testing devices. (A) Testing session for vision-based approaches. (B) Testing session for an angular measurement sensor-based approach.

accurate 3D positions of the markers with an accuracy of up to 0.1 mm. The markers were attached to the subject's center and left body joints, including neck, low back, shoulder, elbow, wrist, hip, knee and ankle joints. Two Optotrak^{TM} cameras were positioned to the left side of the subject to prevent possible data loss due to occlusions of the markers.

In the session for the angular measurement sensor-based approach (Fig. 3B), the optical encoder (550 samples per second) was positioned across the left knee to measure knee-included angles; the optical encoder was placed using a specially designed exoskeleton described in ref. [37], reducing the effect of soft tissue movements. To obtain ground truth angles, active markers were attached to left hip, left knee and left ankle joints. Two Optotrak^{TM} cameras were also positioned to the left side of the subject.

In each session, human motion was simultaneously recorded with these devices. For the synchronization of motion data, the subject was asked to hold a T-pose at the beginning of the recording. Data synchronization was manually performed by identifying the T-pose frame across all measurement techniques.

Testing tasks

To compare the accuracy of motion data for diverse tasks, one male subject simulated three types of tasks as shown in Fig. 4: (1) basic tasks with movements of specific body parts; (2) lifting and placing; and (3) walking. The basic tasks were designed to test the measurement accuracy for simple motions that involve movements of specific body parts. Those include arm-raising to the front and the side, elbow-bending, back-bending, back-twisting and knee-bending, which are also common motions in manual work such as construction. For more dynamic motions involving coordinated

https://doi.org/10.1017/S0263574717000571 Published online by Cambridge University Press

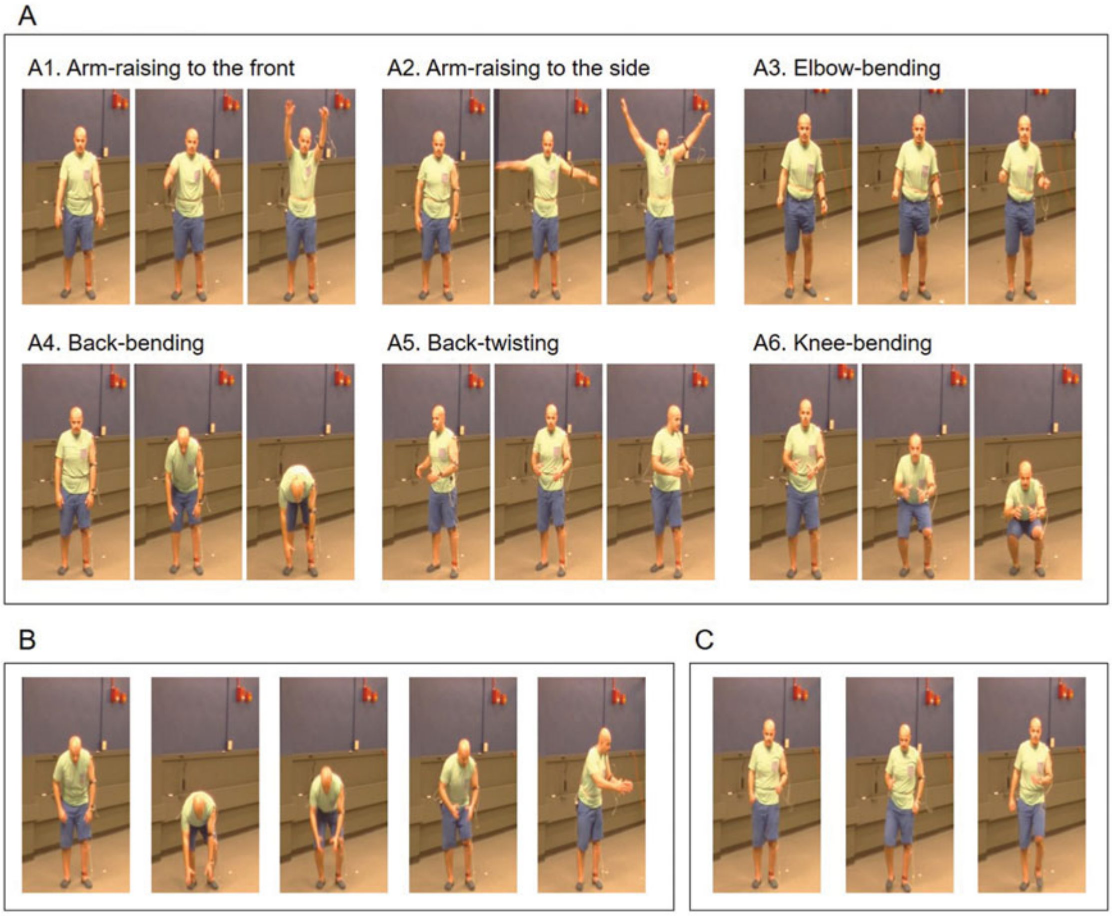

Fig. 4. Testing tasks. (A) Basic tasks. (B) Lifting and placing. (C) Walking.

body movements, a lifting and placing task was selected (also common in construction). Specifically, the subject was asked to simulate the lifting task by lifting an imaginary object from the ground and placing it to the side. Lastly, a walking task was intended to test the measurement accuracy for rapid repetitive movements. To perform identical tasks for two independent sessions, the subject was asked to practice the task in question for several times before recording two test sessions.

Measures for accuracy comparison

As measures of motion data accuracy for vision-based motion capture approaches, previous studies have used 3D positions of body joints, body link lengths or joint rotation angles.[16–19] However, due to the difference in body models used in each vision-based approach, the use of these measures may lead to bias in accuracy comparison. For example, joint locations and corresponding body link lengths in a multiple camera-based approach can be calculated based on the measured joint locations of the subject. On the other hand, the RGB-D sensor- and stereovision camera-based approaches capture motions by matching 3D point clouds with a pre-defined body model, and thus the body link length from the captured motion data is affected by anthropometric mismatch between the model and subject. Also, while both RGB-D sensor- and stereovision camera-based approaches provide motion data in a BVH file format that defines body postures using joint rotational angles, these angles are not available in the motion data from the multiple camera-based approaches used in this test.[16,18]

To address this issue, the authors define new body angles that are available from all vision-based approaches as shown in Fig. 5. Specifically, the body angles of each body part were defined as the angles between the vector of the body segments and the vertical vector. For example, the vector of the upper arm is obtained using 3D shoulder and elbow locations, and the angle between the vector of the upper arm and the vertical vector (y-axis) is calculated as an upper arm (i.e. shoulder) angle.

https://doi.org/10.1017/S0263574717000571 Published online by Cambridge University Press

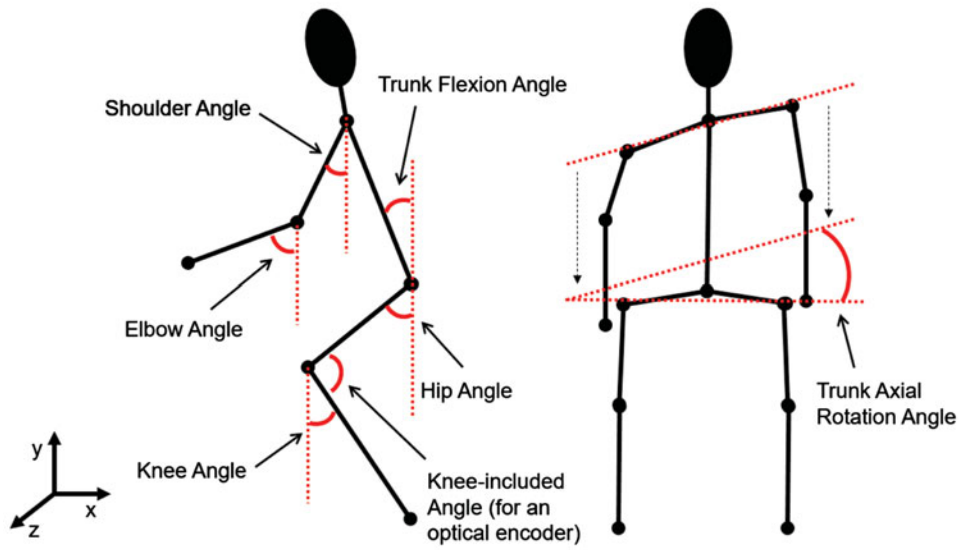

Fig. 5. Body angles to be compared.

The other body angles such as lower arm (i.e., elbow), trunk flexion, upper leg (i.e., hip) and lower leg (i.e., knee) angles are calculated using the same method. However, the trunk axial rotation angle that indicates how much the back is twisted was computed by using shoulder and hip vectors that were projected onto the x–y plane. As the motion data from the three vision-based approaches and Optotrak™ provides 3D locations of body joints, all these angles can be calculated using vectors defined by two selected 3D joint locations, enabling accuracy comparison.

To measure accuracy of body angles from an angular measurement sensor-based approach, knee-included angles directly obtained from an optical encoder were compared with the angles determined by 3D locations of markers attached to hip, knee and ankle joints.

Ground truth body angles were calculated based on 3D marker positions from Optotrak™. The markers were attached to the skin near the joints, not the centers of body joints. As a result, body angles from Optotrak™ may slightly differ from the angles from vision- and angular measurement sensor-based approaches. To adjust possible discrepancies, the body angles were calibrated using the angles from a T-pose. Also, the body angles from each approach were smoothed using a Savitzky–Golay filter[40] that has been widely used for post processing of motion data.[41]

Results

Vision-based motion capture approaches

Figure 6 shows the plots of body angles from the vision-based motion capture approach during one cycle of diverse basic tasks. Through the visual investigation, it was found that overall body angles from each approach were closely matched with body angles from an Optotrak™, while back (flexion and rotation) and upper leg angles from a multiple camera-based approach in particular showed some discrepancies during the middle of the tasks.

For the quantitative assessment during these tasks, mean and standard deviation of absolute errors (MAEs and S.D. of AEs), and maximum and minimum errors (MAX and MIN) in body angles between four different approaches and Optotrak™ were calculated as shown in Table I. The RGB-D sensor-based approach showed the most accurate (4.2 degrees of average MAEs) and reliable (2.8 degrees of average S.D.) results for all body angles. The stereovision camera-based approach also provided relatively accurate motion data, resulting in 6.2 degrees of MAE, but showed higher variations (4.2 degrees of average S.D.) than the RGB-D sensor-based approach. The least accurate results (11.6 degrees of average MAEs) were obtained from the multiple camera-based approach, especially due to relatively larger errors in lower arm (16.2 degrees of MAEs), truck flexion (12.5 degrees of MAEs) and trunk rotation (21.9 degrees of MAEs) angles than other body angles.

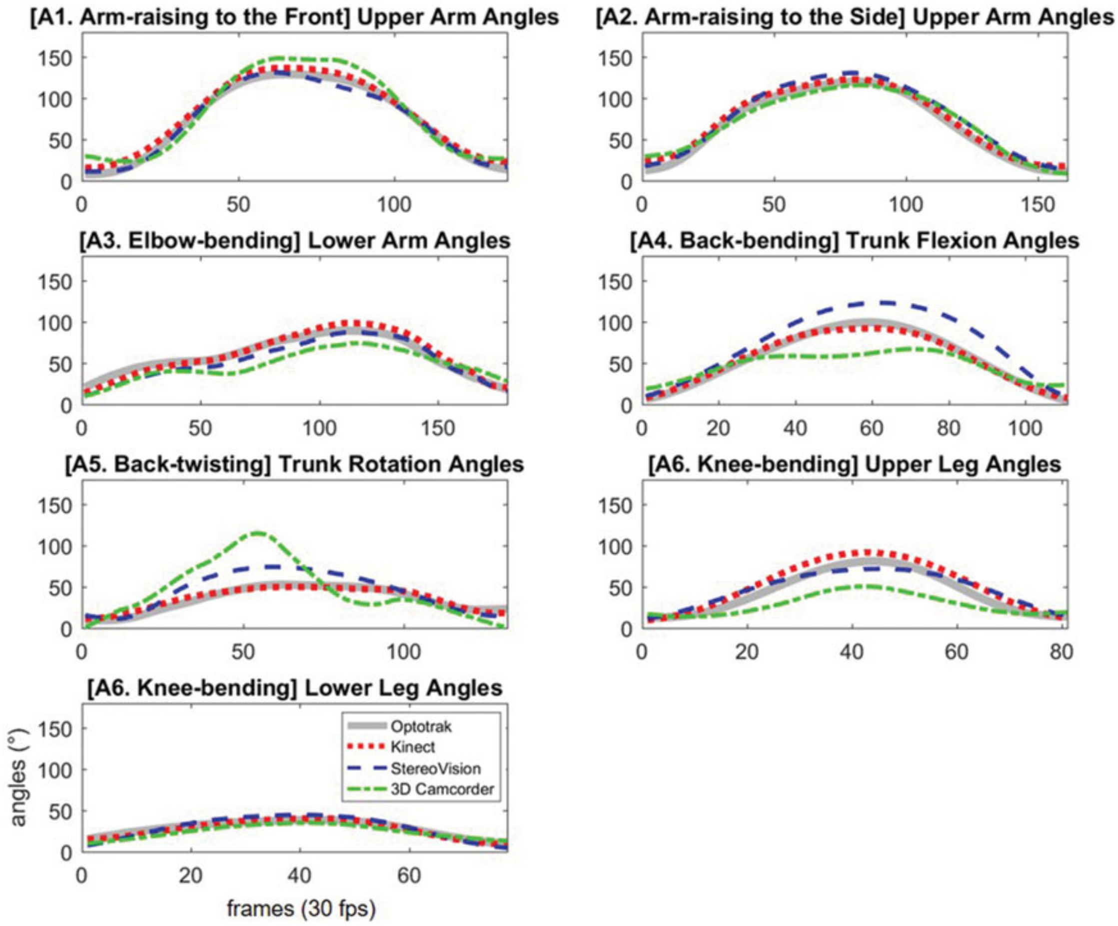

Fig. 6. Comparison of body angles from vision-based approaches during basic tasks.

Figure 7 shows the body angles from vision-based approaches during one cycle of a lifting and placing task. Even for a complex task that involves simultaneous whole body movements, all the approaches provided robust body angle measurements for all body parts. Unlike basic tasks, any severe discrepancies in body angles from a multiple camera-based approach were not observed.

Average MAEs during a lifting and placing task were 6.5 (RGB-D sensor-based), 6.6 (stereovision camera-based) and 10.9 (multiple camera-based) degrees, showing similar errors in body angles during basic tasks (Table II). Both RGB-D sensor- and stereovision camera-based approaches showed robust results in this task, even though the errors in body angles in the RGB-D sensor-based approach were slightly increased. Again, in motion data from the multiple camera-based approach, larger errors in back (torso flexion and rotation) angles were observed, while upper arm angles were relatively accurate.

Last, a walking task showed larger discrepancies in the patterns of body angles from all vision-based approaches as shown in Fig. 8. Motion data from the RGB-D sensor- and multiple camera-based approaches induced similar errors (7.1 and 11.0 degrees of average MAEs, respectively) with other tasks, while the stereovision camera-based approach showed the largest errors (12.6 degrees of average MAEs) among the three tasks (Table III).

An angular measurement sensor-based approach

Figure 9 shows the plots of knee-included angles measured using an optical encoder and an Optotak™ during the three tasks (among basic tasks, only the knee-bending task was tested). Two plots almost matched each other, indicating an accurate angular measurement of an optical encoder for all three tasks. However, small discrepancies were observed at the beginning and end of the cycle of knee-bending and lifting and placing tasks.

https://doi.org/10.1017/S0263574717000571 Published online by Cambridge University Press

Table I. Accuracy of body angles from vision-based approaches during basic tasks (unit: degrees).

Body angles	Metrics	RGB-D sensor (Kinect™)	Stereovision camera (Bumblebee (XB3™)	Multiple camera (3D camcorder)
A1. Arm-raising to the front				
Upper arm	MAE	5.9	3.0	11.3
	S.D. of AE	2.5	2.6	6.8
	MIN/MAX	−9.9 to −0.6	−3.2 to 9.9	−22.0 to 12.0
A2. Arm-raising to the side				
Upper arm	MAE	4.7	8.2	7.6
	S.D. of AE	2.3	3.8	4.2
	MIN/MAX	−11.1 to −1.9	−14.3 to −0.3	−16.8 to 7.6
A3. Elbow-bending				
Lower arm	MAE	4.9	8.1	16.2
	S.D. of AE	3.4	4.0	4.2
	MIN/MAX	−9.8 to 8.6	−2.7 to 14.0	10.1 to 24.5
A4. Back-bending				
Back flexion	MAE	2.5	15.5	12.5
	S.D. of AE	2.1	11.2	12.5
	MIN/MAX	7.6	0.0	39.0
	MIN	−3.9	−34.3	−19.7
A5. Back-twisting				
Back axial rotation	MAE	3.1	11.0	21.9
	S.D. of AE	1.9	8.6	18.5
	MIN/MAX	−6.5 to 4.6	−23.8 to 8.6	−64.6 to 22.5
A6. Knee-bending				
Upper leg	MAE	5.4	4.3	9.8
	S.D. of AE	5.5	4.6	12.0
	MIN/MAX	−13.7 to 3.3	−14.0 to 9.3	−4.9 to 32.4
Lower leg	MAE	1.0	2.4	2.7
	S.D. of AE	1.1	2.6	2.8
	MIN/MAX	−1.7 to 4.1	−6.5 to 7.6	−3.3 to 8.1
Average MAE		4.2	6.2	11.6
Average S.D.		2.8	4.4	8.1

Notes: MAE: mean absolute error; AE: absolute error; S.D.: standard deviation; MAX: maximum value of errors; MIN: minimum value of errors.

The MAE for knee-included angles from an optical encoder was 2.9, 3.8 and 3.0 degrees for the three tasks, respectively (Table IV). Compared with an RGB-D sensor-based approach that showed the most accurate measurements for upper and lower leg angles (1.0−8.8 degrees of MAEs) among vision-based approaches, this approach provided the most accurate and reliable angular measurements regardless of types of tasks.

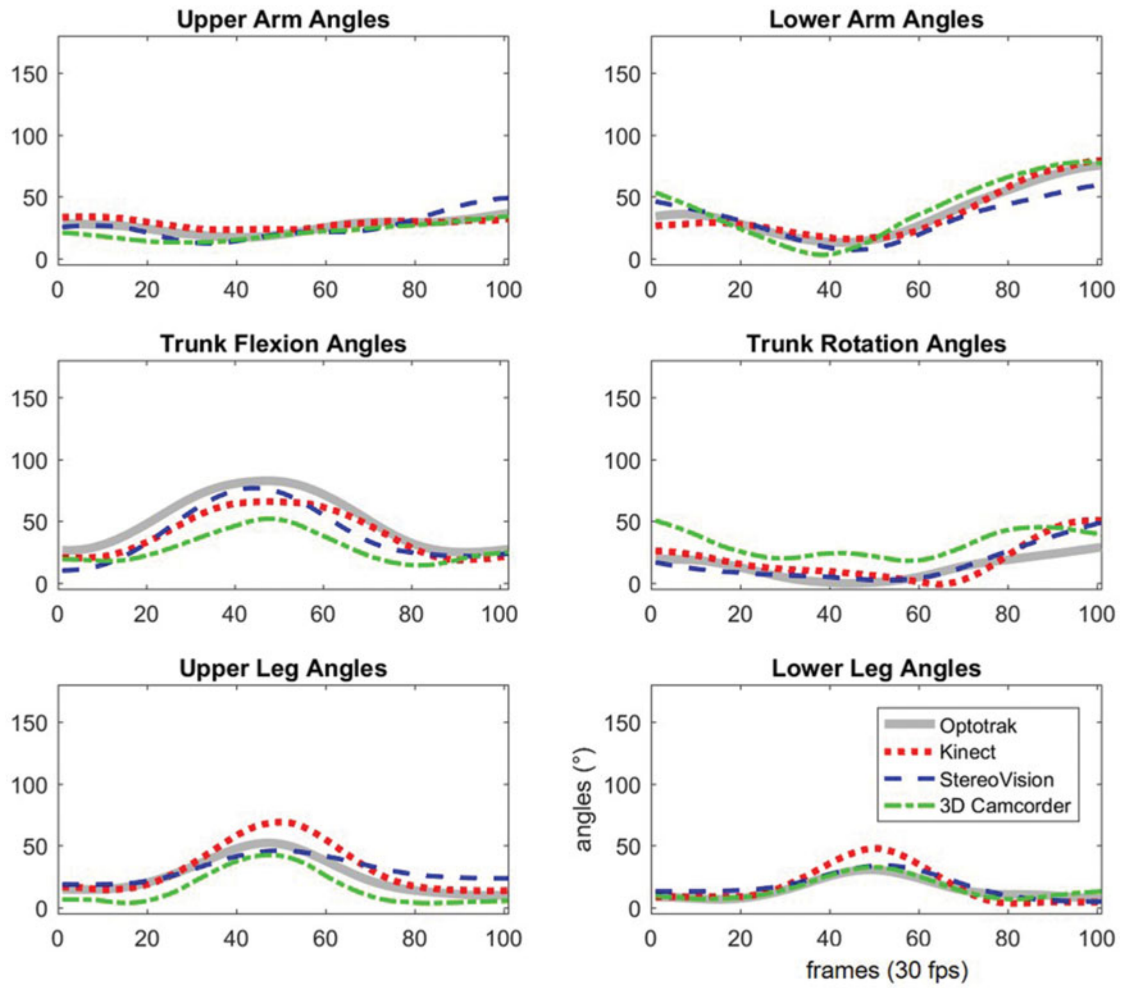

Fig. 7. Comparison of body angles from vision-based approaches during a lifting and placing task.

Discussion

Performance comparison

Specifications and accuracies of three vision-based motion capture approaches and an optical encoder are summarized in Table V. The experimental tests in the previous section presented 5.9, 8.5, 11.2 and 3.2 degrees of average MAEs for RGD-D sensor-based, stereovision camera-based and multiple camera-based approaches and for an optical encoder, respectively. The motion capture performance of each approach tends to rely on the specifications of devices (e.g., types of raw data, resolution and fps). For better decisions on appropriate uses of these approaches in construction, it would be important not only to compare the accuracy but also to understand comparative advantages and limitations.

Among vision-based approaches, the RGB-D sensor-based approach showed the most accurate and reliable results for all three tasks as it uses data-rich 3D images and has a high resolution and frame rate. It is also expected that rapid technological development of RGB-D sensors will enable us to collect more accurate and reliable 3D point cloud data, contributing to improvement of motion tracking performance. Despite the robust performance of this approach, its short operating range (less than 4m) and sensitivity to sunlight may limit its application to confined and indoor areas.

Alternatively, the stereovision camera-based approach can be a practical solution by taking advantage of its ability to collect 3D images at both indoor and outdoor conditions and longer operating range. The accuracy of body angles from this approach was also not much different from the RGB-D sensor-based approach, when excluding the walking task. Considering that walking

Table II. Accuracy of body angles from vision-based approaches during a lifting and placing task (unit: degrees).

Body angles	Metrics	RGB-D sensor (Kinect™)	Stereovision camera (Bumblebee (XB3™)	Multiple camera (3D camcorder)
Upper arm	MAE	3.5	4.6	4.4
	S.D. of AE	2.4	3.8	3.3
	MIN/MAX	−6.3 to 4.4	−13.1 to 6.5	0.3 to 10.4
Lower arm	MAE	3.6	7.6	7.5
	S.D. of AE	1.9	4.7	3.6
	MIN/MAX	−4.9 to 8.3	−12.1 to 16.2	−19.3 to 11.1
Back flexion	MAE	10.3	11.0	22.7
	S.D. of AE	5.1	5.2	11.2
	MIN/MAX	2.9 to 16.9	3.1 to 18.4	2.2 to 35.5
Back axial rotation	MAE	8.4	5.5	18.8
	S.D. of AE	6.2	5.2	4.8
	MIN/MAX	−23.4 to 11.3	−20.0 to 7.2	−31.1 to −10.5
Upper leg	MAE	6.9	7.1	10.3
	S.D. of AE	6.2	4.7	2.7
	MIN/MAX	−19.4 to 1.5	−13.9 to 7.0	4.6 to 14.9
Lower leg	MAE	6.0	4.0	1.5
	S.D. of AE	5.0	1.8	1.4
	MIN/MAX	−17.5 to 7.2	−6.7 to 4.0	−6.1 to 3.7
Average MAE		6.5	6.6	10.9
Average S.D.		4.5	4.2	4.5

Notes: MAE: mean absolute error; AE: absolute error; S.D.: standard deviation; MAX: maximum value of errors; MIN: minimum value of errors.

involves more rapid movements than other tasks in this test, it was likely that the low frame rate (8–10 fps) of the stereovision resulted in tracking errors of certain body parts (e.g., upper limbs) that moved quickly. As the frame rate of a stereovision camera is determined by the computational time for 3D reconstruction and the performance of hardware, the use of an advanced 3D reconstruction algorithm and a high performance computer can achieve a higher frame rate that helps to reduce errors in motion data, particularly during tasks involving rapid body movements. Regarding the operating range of the stereovision camera-based approach, it is recommended to set Bumblebee XB3™ within 10 m as the quality of 3D point clouds is significantly affected by the distance from the scene. However, a binocular stereovision system theoretically works with any type of two 2D cameras that are separated by a short distance, and are mounted parallel to one another. As a result, this approach is flexible in terms of operating ranges if zoom lenses are used. Recently, a stereovision system with adjustable zoom lens control has been introduced,[42] enabling more practical application of this approach.

The multiple camera-based approach showed larger errors in body angles than the other two vision-based approaches. The RGB-D sensor-based and stereovision camera-based approaches benefit from the 3D imaging hardware that provides richer information (e.g., RGB pixel values + depth information) on scenes. However, the multiple camera-based approach need to extract motion data by processing only 2D images that contain less information (e.g., RGB pixel values). An inaccurate camera calibration process could also lead to errors in 3D triangulation of body joints from two images. Considering these limitations, a multiple camera-based approach with about 10 degrees of error in body angles is promising. Despite relatively larger errors, the multiple camera-based approach has several competitive advantages from a practical point of view, compared with the other

https://doi.org/10.1017/S0263574717000571 Published online by Cambridge University Press

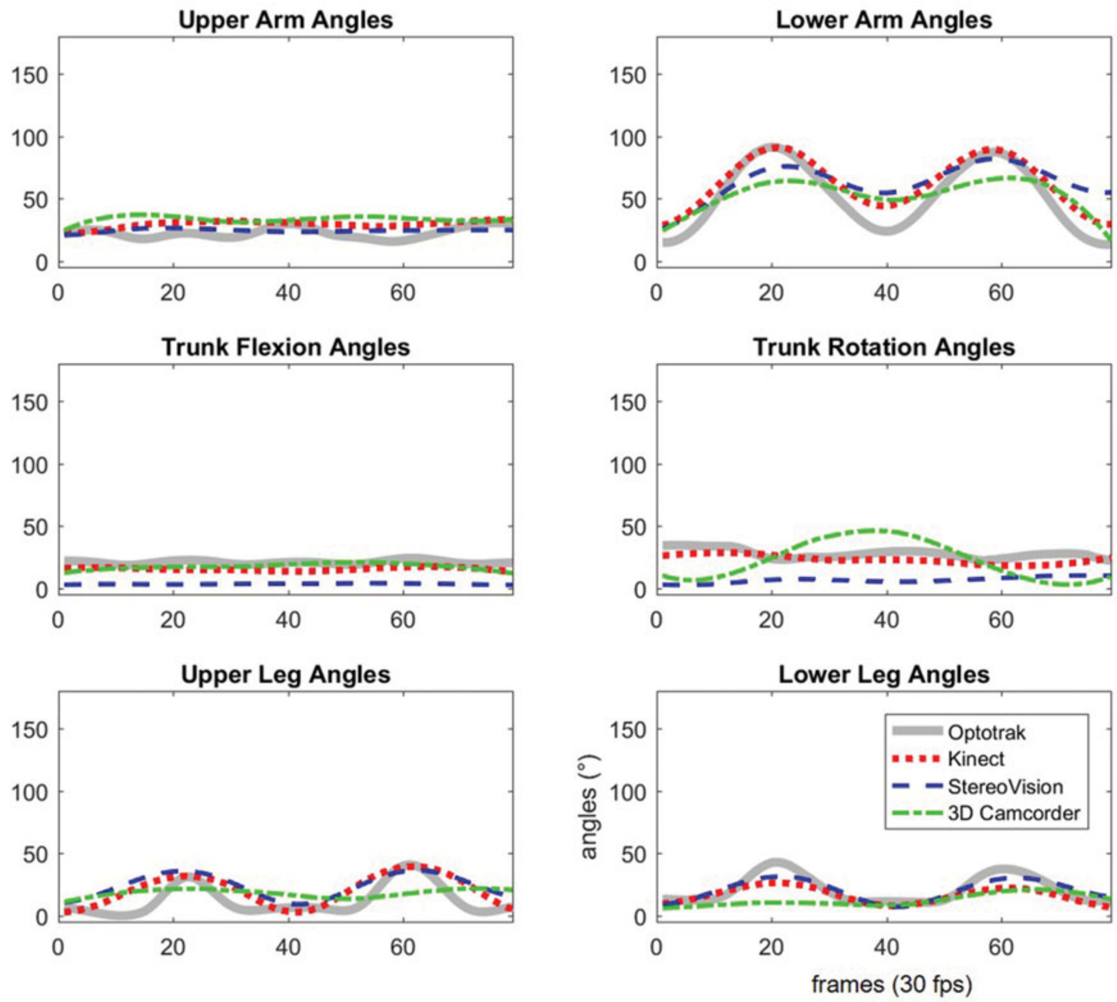

Fig. 8. Comparison of body angles from vision-based approaches during a walking task.

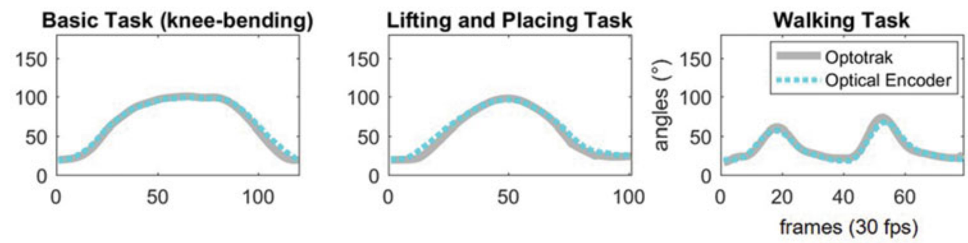

Fig. 9. Comparison of knee-included angles from an angular measurement sensor-based approach (i.e., an optical encoder) during the three tasks.

two approaches. For example, as any types of ordinary cameras can be used to collect 2D images, additional investments in devices are not required. Due to the use of zoom lenses, its operating range is theoretically unlimited. Less sensitivity to rapid movements is another strength of this approach. In addition, there is room for further improvement if occlusion issues are handled. One of the reasons of the least accurate results from the multiple camera-based approach is that it showed relatively larger errors in lower arm, truck flexion and trunk rotation angles than other body angles. As shown in A3 (elbow-bending), A4 (back-bending) and A5 (back-twisting) tasks in Fig. 4, an elbow or a hip was occluded by a lower arm or a torso (i.e., self-occlusions), which may lead to incorrect detections of these joints in the multiple camera-based approach. In these tests, especially, a 3D camcorder was

https://doi.org/10.1017/S0263574717000571 Published online by Cambridge University Press

Table III. Accuracy of vision-based motion capture approaches during a walking task (unit: degrees).

Body angles	Metrics	RGB-D sensor (Kinect™)	Stereovision camera (Bumblebee (XB3™)	Multiple camera (3D camcorder)
Upper arm	MAE	7.1	4.5	10.9
	S.D. of AE	4.0	2.3	5.5
	MIN/MAX	−13.3 to 1.8	−8.7 to 6.5	−19.3 to −2.0
Lower arm	MAE	10.7	15.9	15.0
	S.D. of AE	6.3	11.6	7.7
	MIN/MAX	−20.6 to 1.1	−41.8 to 17.1	−25.4 to 28.0
Back flexion	MAE	5.4	17.3	3.5
	S.D. of AE	1.5	1.4	2.2
	MIN/MAX	2.3 to 7.9	15.0 to 20.4	−1.7 to 9.4
Back axial rotation	MAE	4.8	21.3	15.3
	S.D. of AE	4.8	5.3	8.1
	MIN/MAX	−24.4 to 8.8	14.6 to 32.0	−18.6 to 27.9
Upper leg	MAE	8.8	12.1	11.6
	S.D. of AE	6.6	7.2	5.5
	MIN/MAX	−21.2 to 3.8	−27.3 to 4.9	−18.9 to 23.1
Lower leg	MAE	5.6	4.2	9.7
	S.D. of AE	5.1	2.7	8.9
	MIN/MAX	−4.3 to 16.6	−6.9 to 11.6	−5.0 to 32.5
Average MAE		7.1	12.6	11.0
Average S.D.		4.7	5.9	6.3

Notes: MAE: mean absolute error; AE: absolute error; S.D.: standard deviation; MAX: maximum value of errors; MIN: minimum value of errors.

Tablle IV. Accuracy of an angular measurement sensor-based approach (i.e., an optical encoder) during the three tasks (unit: degrees).

Body angles	Metrics	Basic task	Lifting and placing task	Walking task
Knee-included	MAE	2.9	3.8	3.0
	S.D. of AE	2.7	3.1	2.1
	MIN/MAX	−10.1 to 2.0	−10.7 to 1.8	−5.6 to 8.8

Notes: MAE: mean absolute error; AE: absolute error; S.D.: standard deviation; MAX: maximum value of errors; MIN: minimum value of errors.

used to obtain two images from different views. As the distance between two lenses is very short (3.5 cm), both images are affected by self-occlusions. If two independent cameras are positioned away from each other, it could be possible to obtain at least one clear view of images, reducing errors due to self-occlusions.

The optical encoder provided quite accurate measurements for knee-included angles across all types of tasks. Further, as these sensors are attached to body joints to directly measure joint angles, they can provide robust angular measurements for body joints with one degree of freedom under any condition. Although angular measurement sensors, such as the optical encoder, can be used for all body joints, the use of these sensors could be limited due to the need for straps or exoskeletons that may lead to interfering with on-going work. Instead, using angular measurement sensor-based approaches for selected body joints can offset the limitation of vision-based approaches that are sensitive to self-occlusions. However, soft tissue movements may result in errors in body angles from these sensors. For example, during the testing of this approach, small differences in knee-included

Table V. Comparison of specifications and accuracies of vision-based motion capture approaches.

Performance		RGB-D sensor (Kinect™)	Stereovision camera (Bumblebee XB3™)	Multiple camera (3D camcorder)	Optical encoder
Specifica-tions	Raw data	3D images	3D images	2D images	Body angles
	Operating range	Less than 4 m	Less than 10 m (unlimited, with zoom lenses)	Unlimited with zoom lenses	Unlimited
	Resolution	640×480	320×240	1920×1080	–
	fps	30	8–10	29	550
Accuracy (MAEs)	Basic tasks	4.2°	6.2°	11.6°	2.9°
	Lifting and placing	6.5°	6.6°	10.9°	3.8°
	Walking	7.1°	12.6°	11.0°	3.0°
	Average	5.9°	8.5°	11.2°	3.2°

angles were observed at the beginning and end of cycles, which can be attributed to soft tissue movements, especially when straps are not firmly secured to the leg. Securing straps firmly to the body to hold the sensor in position is an important factor to obtain accurate body angles from the sensor.[43]

Potential application areas of in-field motion capture approaches in construction

Vision- and angular measurement sensor-based motion capture approaches tested in this study are considered practical means of in-field motion capture, even though about 5–10 degrees of error in body angles from vision-based approaches and about 3 degrees of error from an angular measurement sensor-based approach still exist. In construction, tasks are performed in unstructured and varying environments, and thus work methods and postures are changing over time. Collecting motion data using these non-invasive and cost effective approaches enables us to understand how workers interact with the environment at construction sites and to identify potential safety and health risks under given environments, specifically when accuracies would not significantly matter such as rough postural assessment, time and motion study and trajectory analysis.

For example, these approaches can be used to specify the severity of working postures. Existing postural ergonomic risk assessment methods determine the level of ergonomic risks based on classified postures through human observation.[44] Some methods such as Rapid Upper Limb Assessment (RULA)[45] and Rapid Entire Body Assessment (REBA)[46] require detailed segmentations of body postures according to body angles. For example, in RULA, trunk postures are categorized into four groups according to trunk flexion angles (0°, 0–20°, 20–60° and over 60°). Body angles obtained from these approaches can be used for rough posture classification that is needed for postural risk assessments. Also, as continuously measured workers' motions during performing tasks is enabled, diverse in-depth motion analysis for understanding physical demands can be facilitated. Traditionally, the pre-determined-motion-time-systems have been widely used to identify workloads during occupational tasks.[47] As these systems rely on human observations to describe workers' manual activities, significant human efforts are generally required. However, by using a time series motion data from the presented approaches, it is possible to accurately and automatically quantify motion-time values for these systems. In addition, trajectory analysis through in-field motion measurements helps to evaluate work efficiency, as well as the risk of ergonomic injuries. For example, shorter trajectories of body movements may imply more efficient movements of a human body, indicating smaller physical demands. Previous studies on movement patterns during occupational tasks found that a more "dynamic" pattern of movements is believed to be associated with a lower incidence of WMSD development.[48,49] Analysis of motion patterns and trajectories using vision-based motion data can broaden our understanding on the job from an ergonomic perspective.

In-field kinematic measurement using vision- and angular measurement sensor-based motion capture approaches has also great potential to be used for more in-depth analysis of physical demands such as biomechanical analysis, even though further improvement of motion data accuracy is required.[50] Biomechanical analysis aims to estimate musculoskeletal stresses as a function of motion and external force data.[4,51] Previous biomechanical studies have relied on laboratory experiments to collect motion data using marker-based motion capture approaches, which can be replaced by in-field motion capture approaches that enable on-site biomechanical analysis. As accurate measurement of all joint angles is necessary for reliable biomechanical analysis, further accuracy improvement of vision-based motion capture approaches is required. However, the sensitivity of biomechanical analysis results to motion errors varies depending on body joints.[52] For example, Chaffin and Erig[52] found that an error of ± 10 degrees in the limiting joint angles (e.g., knees and ankles) could cause the biomechanical analysis results to vary up to $\pm 12\%$ during lifting, pushing and pulling tasks, whereas errors in other joints could have little or no effect. This result indicates that some angular errors in body joints that do not involve forceful exertions are acceptable for biomechanical analysis while it is important to obtain accurate body angles for stressful body joints. So, complemented by the relatively accurate angular measurement sensors such as optical encoders that are applied to the limiting joints, the vision-based motion capture approaches enable researchers to perform biomechanical analysis without significantly sacrificing the reliability of biomechanical analysis.

Conclusions

The study describes the potential of vision-based and angular measurement sensor-based approaches as a means of measuring workers' motions. These approaches are compared through laboratory tests while performing three different types of tasks. Especially, the accuracy of these approaches was computed by comparing body angles from each approach and a marker-based motion capture. The comparison results indicated that the overall errors in body angles from vision-based approaches are about more or less 10 degrees, while an optical encoder, that is one example of angular measurement sensors, can provide quite accurate body angle measurements (about 3 degrees) for specific body joints with one degree of freedom. Self-occlusions and rapid movements are major factors that lead to errors in vision-based approaches.

From a practical perspective, vision-based and angular measurement sensor-based approaches have great potential as non-invasive motion data collection methods at construction sites. Even though several obstacles such a limited operating range (RGB-D sensor-based), low frame rates (stereovision camera) and occlusions (multiple camera-based) still remain to obtain more accurate data from these approaches, further algorithm refinements and hardware developments are expected to address these issues. An angular measurement sensor-based approach such as an optical encoder can provide robust measurements of specific joint movements, despite a small possibility of discomfort by attached sensors. Especially, combined with vision-based motion capture approaches, an angular measurement sensor-based approach can enhance the accuracy of in-field motion measurements. Motion data from these approaches can be used for diverse in-depth analysis without sacrificing its reliability to better understand workers' physical demands during occupational tasks including construction. Also, understanding of how workers behave under given working environments through kinematics measurements and analysis helps to ergonomically design automated machines and assistive robots, aiming to both reduce physical demands and enhance workers' capabilities.

Acknowledgments

The work presented in this paper was supported with a National Science Foundation Award (No. IIP 1640633) in the United States, and was partially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. PolyU 25210917). Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation and the Research Grants Council.

References

1. J. G. Everett, "Overexertion injuries in construction," *J. Constr. Eng. Manag.* **125**(2), 109–114 (1999).
2. Center for Construction Research and Training (CPWR), *The Construction Chart Book: The U.S. Construction Industry and Its Workers* (CPWR Publications, Silver Spring, MD, USA, 2013).

3. V. M. Zatsiorski, *Kinematics of Human Motion* (Human Kinetics, Champaign, IL, USA, 2002).
4. R. G. Radwin, W. S. Marras and S. A. Lavender, "Biomechanical aspects of work-related musculoskeletal disorders," *Theor. Issues Ergon.* **2**(2), 153–217 (2001).
5. T. J. Armstrong, P. Buckle, L. J. Fine, M. Hagberg, B. Jonsson, A. Kilbom, I. A. Kuorinka, B. A. Silverstein, G. Sjogaard and E. R. Viikari-Juntura, "A conceptual model for work-related neck and upper-limb musculoskeletal disorders," *Scand. J. Work Environ. Health*, **19**(2), 73–84 (1993).
6. J. G. Everett and A. H. Slocum, "Automation and robotics opportunities: Construction versus manufacturing," *J. Constr. Eng. Manag.* **120**(2), 443–452 (1994).
7. A. Schiele and F. C. van der Helm, "Kinematic design to improve ergonomics in human machine interaction," *IEEE Trans. Neural Syst. Rehabil. Eng.* **14**(4), 456–469 (2006).
8. K. Aminian and B. Najafi, "Capturing human motion using body-fixed sensors: Outdoor measurement and clinical applications," *Comput. Animat. Virtual Worlds*, **15**(2), 79–94 (2004).
9. T. Takubo, Y. Imada, K. Ohara, Y. Mae and T. Arai, "Rough Terrain Walking for Bipedal Robot by Using ZMP Criteria Map," *Proceedings of the IEEE International Conference on Robotics and Automation (May 12–17, 2009)* pp. 788–793.
10. S. R. Cruz-Ramírez, Y. Mae, T. Arai, T. Takubo and K. Ohara, "Vision-based hierarchical recognition for dismantling robot applied to interior renewal of buildings," *Comput. Aided Civ. Inf.* **26**(5), 336–355 (2011).
11. T. Jacobs, U. Reiser, M. Hägele and A. Verl, "Development of Validation Methods for the Safety of Mobile Service Robots with Manipulator," *Proceedings of the 7th German Conference on Robotics ROBOTIK2012*, Munich, Germany (May 21–22, 2012) pp. 117–122.
12. G. Migliaccio, J. Teizer, T. Cheng and U. Gatti, "Automatic Identification of Unsafe Bending Behavior of Construction Workers Using Real-Time Location Sensing and Physiological Status Monitoring," *Proceedings of the Construction Research Congress CRC2012*, West Lafayette, IN, USA (May 21–23, 2012).
13. S. J. Ray and J. Teizer, "Real-time construction worker posture analysis for ergonomics training," *Adv. Eng. Inform.* **26**(2), 439–455 (2012).
14. C. Georgoulas, T. Linner and T. Bock, "Towards a vision controlled robotic home environment," *Automat. Constr.* **39**, 106–116 (2014).
15. T. Linner, J. Güttler, T. Bock and C. Georgoulas, "Assistive robotic micro-rooms for independent living," *Automat. Constr.* **51**, 8–22 (2015).
16. S. Han and S. Lee, "A vision-based motion capture and recognition framework for behavior-based safety management," *Automat. Constr.* **35**, 131–141 (2013).
17. S. Han, M. Achar, S. Lee and F. Peña-Mora, "Empirical assessment of a RGB-D sensor on motion capture and action recognition for construction worker monitoring," *Vis. Eng.* **1**(1), 1–13 (2013).
18. M. Liu, S. Han and S. Lee, "Tracking-based 3D human skeleton extraction from stereo video camera toward an on-site safety and ergonomic analysis," *Constr. Innov.* **16**(3), 348–367 (2016).
19. R. Starbuck, J. Seo, S. Han and S. Lee, "A Stereo Vision-Based Approach to Marker-Less Motion Capture for On-Site Kinematic Modeling of Construction Worker Tasks," *Proceedings of the 15th International Conference on Computing in Civil and Building Engineering ICCCBE2014*, Orlando, FL, USA (Jun. 23–25, 2014).
20. A. A. Alwasel, K. Elrayes, E. M. Abdel-Rahman and C. T. Haas, "A Human Body Posture Sensor for Monitoring and Diagnosing MSD Risk Factors," *Proceedings of the 30th International Symposium on Automation and Robotics in Construction ISARC2013*, Montreal, Canada (Aug. 11–15, 2013).
21. R. Poppe, "A survey on vision-based human action recognition," *Image Vis. Comput.* **28**(6), 976–990 (2010).
22. M. W. Lee and I. Cohen, "A model-based approach for estimating human 3D poses in static images," *IEEE Trans. Pattern Anal.* **28**(6), 905–916 (2006).
23. C. Plagemann, V. Ganapathi, D. Koller and S. Thrun, "Real-Time Identification and Localization of Body Parts from Depth Images," *Proceedings of the IEEE International Conference on Robotics and Automation ICRA2010*, Anchorage, AK, USA pp. 3108–3113.
24. M. Siddiqui and G. Medioni, "Human Pose Estimation from a Single View Point, Real-Time Range Sensor," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops CVPRW2010*, San Francisco, CA, USA (Jun. 13–18, 2010) pp. 1–8.
25. J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake and R. Moore, "Real-time human pose recognition in parts from single depth images." *Commun. ACM*, **56**(1), 116–124 (2013).
26. N. Rafibakhsh, J. Gong, M. K. Siddiqui, C. Gordon and H. F. Lee, "Analysis of Xbox Kinect Sensor Data for Use on Construction Sites: Depth Accuracy and Sensor Interference Assessment," *Proceedings of Constitution Research Congress CRC2012*, West Lafayette, IN, USA (May 21–23, 2012) pp. 848–857.
27. Z. Zhang, "Microsoft kinect sensor and its effect," *IEEE Multimed. Mag.* **19**(2), 4–10 (2012).
28. S. Jin, J. Cho, X. D. Pham, K. M. Lee, S. K. Park, M. Kim and J. W. Jeon, "FPGA design and implementation of a real-time stereo vision system," *IEEE Trans. Circ. Syst. Vid.* **20**(1), 15–26 (2010).
29. J. I. Woodfill, G. Gordon and R. Buck, "Tyzx Deepsea High Speed Stereo Vision System," *Proceedings of the Conference on Computer Vision and Pattern Recognition Workshop CVPRW2004*, Washington, DC, USA (Jun. 27–Jul. 2, 2004).
30. C. Shan, T. Tan and Y. Wei, "Real-time hand tracking using a mean shift embedded particle filter," *Pattern Recogn.* **40**(7), 1958–1970 (2007).
31. C. Liu, J. Yuen and A. Torralba, "SIFT flow: Dense correspondence across scenes and its applications," *IEEE Trans. Pattern Anal.* **33**(5), 978–994 (2011).

32. J. R. R. Uijlings, A. W. M. Smeulders and R. J. H. Scha, "Real-time visual concept classification," *IEEE Trans. Multimed.* **12**(7), 665–681 (2010).
33. H. Bay, T. Tuytelaars and L. van Gool, "SURF: Speeded up robust features," *Comput. Vis. Image Underst.* **110**(3), 346–359 (2008).
34. R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision* (Cambridge University Press, Cambridge, UK, 2003).
35. Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(11), 1330–1334 (2000).
36. M. Tousignant, L. de Bellefeuille, S. O'Donoughue and S. Grahovac, "Criterion validity of the cervical range of motion (CROM) goniometer for cervical flexion and extension," *Spine* **25**(3), 324–330 (2000).
37. A. A. Alwasel, E. M. Abdel-Rahman and C. T. Haas, "A Technique to Detect Fatigue in the Lower Limbs," *Proceedings of the ASME 2014 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, New York, USA (Aug. 17–20, 2014).
38. T. Yamada, Y. Hayamizu, Y. Yamamoto, Y. Yomogida, A. Izadi-Najafabadi, D. N. Futaba and K. Hata, "A stretchable carbon nanotube strain sensor for human-motion detection," *Nat. Nanotechnol.* **6**(5), 296–301 (2011).
39. P. H. Veltink, H. B. Bussmann, W. De Vries, W. L. Martens and R. C. Van Lummel, "Detection of static and dynamic activities using uniaxial accelerometers," *IEEE Trans. Neural Syst. Rehabil. Eng.* **4**(4), 375–385 (1996).
40. A. Savitzky and M. J. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *J. Anal. Chem.* **36**(8), 1627–1639 (1964).
41. P. Esser, H. Dawes, J. Collett and K. Howells, "IMU: Inertial sensing of vertical CoM movement," *J. Biomech.* **42**(10), 1578–1581 (2009).
42. M. Y. Kim, S. M. Ayaz, J. Park and Y. Roh, "Adaptive 3D sensing system based on variable magnification using stereo vision and structured light," *Opt. Laser Eng.* **55**, 113–127 (2014).
43. P. J. Rowe, C. M. Myles, S. J. Hillmann and M. E. Hazlewood, "Validation of flexible electrogoniometry as a measure of joint kinematics," *J. Physiother.* **87**(9), 479–488 (2001).
44. G. Li and P. Buckle, "Current techniques for assessing physical exposure to work-related musculoskeletal risks, with emphasis on posture-based methods," *Ergonomics* **42**(5), 674–695 (1999).
45. L. McAtamney and E. N. Corlett, "RULA: A survey method for the investigation of work-related upper limb disorders," *Appl. Ergon.* **24**(2), 91–99 (1993).
46. S. Hignett and L. McAtamney, "Rapid entire body assessment (REBA)," *Appl. Ergon.* **31**(2), 201–205 (2000).
47. W. Laurig, F. M. Kühn and K. C. Schoo, "An approach to assessing motor workload in assembly tasks by the use of predetermined-motion-time systems," *Appl. Ergon.* **16**(2), 119–125 (1985).
48. Å. Kilbom and J. Persson, "Work technique and its consequences for musculoskeletal disorders," *Ergonomics*, **30**(2), 273–279 (1987).
49. Å. Kilbom, "Repetitive work of the upper extremity: Part II-the scientific basis (knowledge base) for the guide," *Int. J. Ind. Ergon.* **14**(1), 59–86 (1994).
50. J. Seo, R. Starbuck, S. Han, S. Lee and T. Armstrong, "Motion-data–driven biomechanical analysis during construction tasks on sites," *J. Comput. Civil Eng.* **29**(4), (2014).
51. D. B. Chaffin, G. Andersson and B. J. Martin, *Occupational Biomechanics*, 4th ed. (Wiley, New York, USA, 2006).
52. D. B. Chaffin and M. Erig, "Three-dimensional biomechanical static strength prediction model sensitivity to postural and anthropometric inaccuracies," *IIE Trans.* **23**(3), 215–227 (1991).
53. S. Knoop, S. Vacek and R. Dillmann, " Sensor Fusion for 3D Human Body Tracking with an Articulated 3D Body Model," *Proceedings of the IEEE International Conference on Robotics and Automation ICRA2006* pp. 1686–1691.