An Integrated Sensor Network Method for Safety Management of Construction Workers

Tingsong Chen[a], Nobuyoshi Yabuki[b] and Tomohiro Fukuda[c]

[a,b,c]Graduate School of Engineering, Osaka University, Japan
E-mail: chen.tingsong@it.see.eng.osaka-u.ac.jp, yabuki@see.eng.osaka-u.ac.jp, fukuda@see.eng.osaka-u.ac.jp

Abstract –

With the development of the construction industry, many problems such as human safety are remaining unsolved. The construction holds the worst safety record compared to other industrial sectors, approximately 88% of accidents are related to workers' safety. The high complexity of the construction site compare to the ordinary living environment is also a major factor that cannot effectively protect the safety of workers. In this paper, an integrated sensor network method is proposed for the safety management of construction workers. The main signals collected in this paper are visual signals and electronic signals. The compatibility issues caused by cooperation between different types of signals will also be discussed in this paper. At the same time, a multi-signal automatic correction method is used to improve the accuracy and efficiency of our proposed method.

Keywords –

sensor fusion; motion recognition; safety management; IMU; depth camera.

1 Introduction

"Smart City" concept is widely known by people around the world for many years, the main purpose of it is using information technology to help to improve city service. Currently, Smart City is applied in transportation, citizen management, urban resource allocation and so on, they all perform well compare to traditional methods. Along with the rising of Smart City, the concept of "Smart Construction" is also proposed recently, many well applied methods from other fields are poured into construction area. But due to the differences in management mode and implementation, those methods perform not so well. There is a lot of room for development of Smart Construction.

Speaking of which, construction industry field still holds a worst record in safety compare to others. In Japan, death number in construction accidents is twice that of Germany and 3 times that of UK from data since 2003 to 2005 [1], and haven't improve well during the last 20 years [2]. Comparing to other developed countries, accident monitoring efficiency is the main reason that workers cannot be found and rescue in time.

The main causes of workers' injuries and deaths are heat stroke, hitting by heavy objects, felling from a height and so on. Some of the injuries are caused by accident, some of them happened because of the unsupervised unsafe acts of workers' own due to cost, time pressure and other reasons. Normal monitoring system such as web camera requires manual operation, it is inefficient because of human neglect, obstacle and other factors. A more efficient and accurate safety management and monitoring method is needed.

Camera-based monitoring method is widely used and researched around the world, it has a lot of benefits, such as low cost, easy to be assembled, and so on. Yet more disadvantages are unavoidable, such as unable to solve occlusion problem, low accuracy in weak light or dark environment. Other methods concentrate on changing RGB frame to RGB-D frame by adding depth into the picture, such as Kinect and Realsense. RGB-D is more accurate to detect human and object comparing to RGB's pixel crop, it also works well even in weak light environment. However, due to the limit of working distance, reflective surfaces and relative surface angles, depth maps in RGB-D frame always contain significant holes and serve noise, as shown in Figure 1, these errors limit the practical usage of RGB-D frame in real applications, thus depth maps restoration in hole filling and noise removing becomes a necessary step in depth-camera-based monitoring system.

Monitoring method based on Inertial Measurement Unit (IMU) sensors is also a hot topic in recent years, because of its undoubtable benefits compare to other methods such as visual camera. IMU sensors are nonintrusive, lightweight and portable measuring devices, they can overcome the sensor viewpoint and occlusion issues, once they were attached on subject, the activities can be detected in a non-hindering manner [4], [5], [6]. After pre-processing of the motion for activity

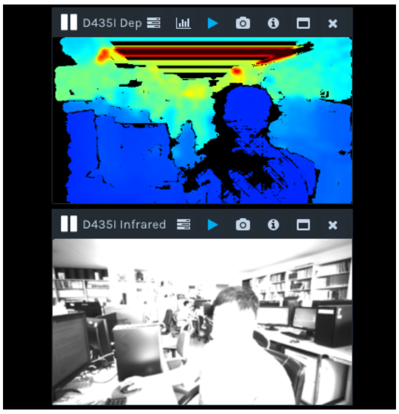

Figure 1. Depth map (top) and infrared frame

recognition, discriminative features are then derived from time and/or frequency domain representation of the motion signals [7] and used for activity classification [8]. Although there are a lot of benefits of IMU sensors, there are also some disadvantages, first, it is not intuitive, not easy for manual recheck, second, model complexity is hard to control especially when precise motion capture is needed.

In order to fix injuries problems in construction field, we propose a safety management method based on multiple sensor network. This method aims at using data of workers from different kinds of sensors to detect human motion and movement more accurately, sensors such as depth camera and IMU sensor will be used, and by cooperating them to decrease the error rate and improve efficiency in accident alert and injury rescue.

The main contributions of this paper are as below:

1. Improve the accuracy of multi-type sensor fusion based motion recognition in specific motions in construction site environment;
2. Multi-type sensor fusion based motion recognition correction method.

The remaining sections are organized as follows. Related works are reviewed in Section 2, including camera-based, depth camera-based and IMUs-based methods. The proposed methodology is introduced in Section 3, visual sensor-based human recognition, IMU s-based human recognition and sensor fusion method will be introduced. The simulation and experiment details and hardware parameters will be introduced in Section 4. The conclusion is drawn in Section 5.

2 Related works

2.1 Camera-based human modeling

Commercial camera-based human detection method requires subjects wear markers and depend on multiple calibrated cameras mounted in the environment. [9], it is inconvenient, to overcome these constraints, other researchers focus on developing marker-less approaches from multiple cameras, yet some of these methods require offline processing to achieve high quality results [10], [11]. But some other real-time approaches have been proposed [12], these approaches typically fit a skeletal model to image data. Other approaches to real-time performance include combining discriminative and generative approaches [13]. However, multi-view approaches assume stationary and well calibrated cameras, therefore they are not suitable in mobile scenarios.

2.2 Depth camera-based depth map restoration and human modeling

As for regular camera, RGB-based depth prediction normally uses large body of literature, training exclusively using ground truth metric depth, [14,15,16,17,18]. As for depth camera, many methods have been proposed for restoring depth maps by Kinect, these methods can be classified into two types: filtering-based and reconstruction-based. Filtering-based methods use different filters to restore captured depth maps. Lai et al. [19] applied a median filter in RGB space to fill holes in depth map recursively, however this method will blur sharp edges obviously. To preserve sharp edges, Camprani et al. [20] applied a joint bilateral filter in depth map iteratively. Matyunin et al. [21] considered using temporal information to restore depth map, but this method occurs delay because it uses multiple consecutive frames to restore target depth map. Reconstruction-based methods use image inpainting techniques to fix missing values in depth maps. Telea [22] proposed FMM (Fast Marching Method) for image inpainting. Miao et al. [23] proposed a texture-assisted method in which the texture edge information is extracted for assisting depth restoration. These methods can remove noise and fill small holes in depth maps, however when it comes to large holes exist in depth map, such as holes in Figure 1, the results are unsatisfactory.

About human modeling, Anguelov et al. [24] introduced SCAPE, a data-driven method for building a human 3D model than spans variation in both shape and pose. It shows that given a high-resolution range image from a single view, the SCAPE model can be used to observe data. Based on SCAPE parameterized model, Weiss et al. [25] combined multiple views of person and several low-resolution scans to obtain an accurate human 3D model. Liao et al. [26] introduced prior of human body pose and shape, then proposed a human 3D modeling method based on a monocular depth camera.

2.3 IMUs-based human modeling

Roetenberg et al. [27] used 17 IMUs equipped with 3D accelerometers, gyroscopes and magnetometers, fused together using a Kalman Filter. Assuming the measurement are noise-free and contain no drift, the 17 IMU orientations completely define the full pose of the subject (using standard skeletal models). However, 17 IMUs are very intrusive for the subject, long setup times are required, errors such as placing a sensor on the wrong limb are common, which makes it difficult to reproduce. Marcard et al. [28] compute accurate 3D poses using only 6 IMUs. They take a generative approach and place synthetic IMUs on Skinned Multi-Person Linear Model (SMPL) body model [29]. They solve for the sequence of SMPL poses that match the observed sequence of real measurements by optimizing over the entire sequence. But this method relies on computationally expensive offline optimization, which is also hard to reproduce. Therefore, a smaller number of IMU sensors and less computation complexity is the key of future IMUs-based human modeling and motion detection.

3 Methodology

In this paper, we propose an integrated sensor network method for safety management of construction workers. This method mainly uses depth camera, IMU sensor, and environment sensor to collect data from workers and construct human model to analyze human motion, gesture and some physical index such as temperature and air pressure. This method concentrates on multiple sensor cooperation, by using different kinds of sensor to decrease the errors caused by sensor defects, to increase the accuracy of detecting and improve efficiency.

3.1 Depth camera-based human recognition

Yin et al. [3] proposed a two-stage stacked hourglass network based on Varol et al. [30] to get high-quality result of human depth prediction. Instead of using RGB image directly, this method uses RGB image and human part-segmentation together to predict human depth. It consists of convolution layer, part-segmentation module, and depth prediction module. First, RGB image input goes through the convolution layer and turns into heat maps, then enter the part-segmentation module, after then, heat maps turn into human part-segmentation results, these heat maps are summed as the input of the following depth prediction module with previous layers features, finally human depth prediction results are outputted.

Algorithm 1 GradientFMM

```
Procedure GradientFMM (depthmap)
    Known ← all pixels with known values in depthmap
    Unknown ← all unknown pixels adjacent to Known in depthmap
    insert all pixels in Unknown into min-heap
    while Unknown not empty do
        p ← root of min-heap
        calculate p's value using depth value equation
        add p to Known
        remove p from Unknown
        perform down heap
        for each neighbor q of A do
            if q not in Known and Unknown then
                add q to Unknown
                perform up heap
            end if
        end for
    end while
    return Known
end procedure
```

The algorithm above is called GradientFMM [3], it propagates depth from known pixels to unknown pixels. After the process, every pixel in depth map in the unknown region has a depth value. In order to mark useful pixels in depth maps to predict possible human skeleton, we use GradientFMM as our pre-treatment method in human depth prediction.

The resolution of collected image from depth camera is 848x480 and 30 frames per second. In our method, firstly we apply GradientFMM algorithm to analyze each frame and get data of human depth maps, next, we consider frame platform and depth direction as a 3D coordinate, and collect all coordinate data of each pixel inside depth map. At the same time, well-trained image processing algorithm will be used to identify skeleton based on depth map, in this research, we use OpenPose or Intel skeleton tracking SDK. After human skeleton is detected, the 3D coordinate changes of specific parts of human (head, hand, foot) and required parts (arm, waist) will be recorded, and compare with collected database, to find out the best match and output. Frames with skeleton will be used as input for further image processing to improve accuracy of skeleton mapping.

3.2 IMUs-based human recognition

This part introduces IMUs-based human motion detection, IMUs can measure triaxial (3D) accelerations and triaxial angular velocities. It is also easy to obtain information directly without numerous restrictions.

In our proposal, we mainly consider the motion capture while workers are working, so the upper body will be observation focus. 3 IMU sensors will be used to detect movement changes, two will be attached on outer arms, another one will be attached on front waist.

Not only can IMU sensors collect movement data, but also, they are able to collect workers' surrounding environmental factors such as temperature, height, air pressure and so on, by doing data exchange with environment sensors, to ensure that workers are in proper working environment.

IMU motion recognition is based on Dehzangi et al. [31], this paper introduced a human activity recognition method in normal environment, the activities they considered are walking, walking upstairs, walking downstairs, sitting, standing and sleeping. In our proposal, due to the difference of subjects, new motions are added: uplift (one or two arms up), pick up heavy object (the swing amplitude of both arms is reduced and stiff), hold up heavy object (the arms are partially angled and stiff), arms raised (arms at right angles to body), regular cyclical movement (arms turn the roulette), bend over (lean forward or backward).

The framework of IMU-based human activity recognition system is: first, collecting relevant data through users; then, in the learning phase, relevant features are extracted from the time-series raw data. Model complexity is reduced by applying Feature selection/ Dimension reduction technique. Recognition model is created from the dataset of selected features. Then, in the testing phase, this model is used to evaluate raw signal and create an activity label. Finally, raw data will be outputted as labeled motion.

IMU sensors attached in 3 parts will continuously record data, and using motion recognition method to analyze amplitude changes. At the same time, when there is a dramatic change during the process, the differences between triaxial accelerations and angular velocities before and after the change will be counted and recorded as change graphs. Finally, the differences will be compared with motion database and find out the best match.

3.3 Multi-type sensor fusion and analysis

Normally, visual signal and IMU electronic signal are quite different, it is hard to make a comparison between them. In our proposal, both visual camera-based method and IMU sensor-based method can perform results individually, but when it comes to some special occasion such as partly occlusion, using only one kind of signal will cause high error and effect the whole system.

In this research, we try to cooperate two kinds of signal, by educe the advantages and disadvantages of them to further improve accuracy of this method. We consider the whole area as a huge 3D coordinate system, as shown in Figure 3, depth camera is placed on one side of the system, IMU sensors are also calibrated before loading to make sure they are consistent at time 0.

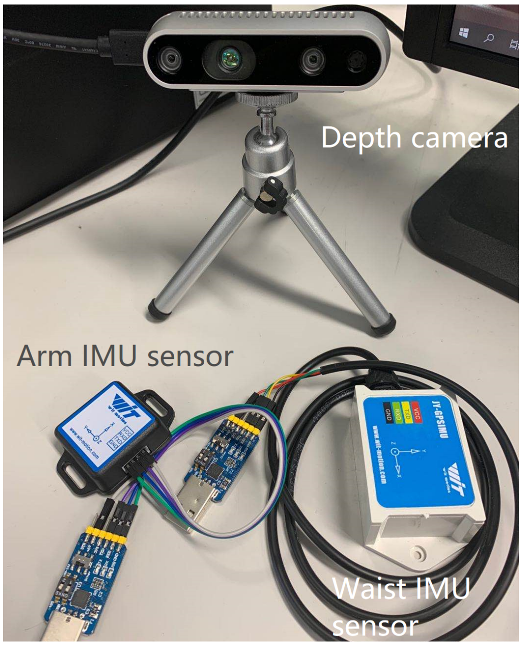

Figure 2. Depth camera and IMU sensors

As introduced in previous sections, when record starts, both camera and IMU sides will generate constant 3D coordinate changes. As for depth camera side, the variation and value of specific point can get from the coordinate made by depth map and frame platform. As for IMU sensor side, during the movement, three axes will change in different accelerations, based on the origin set at time 0, the path changes and distance can be calculated by double integral:

$$\vec{S} = \int(\int(\vec{a})\mathrm{dt})\mathrm{dt} \tag{1}$$

where $\vec{S}$ represent directional distance, $\vec{a}$ represent average acceleration during time period t.

Although the unit, distance and size are quite different between depth map coordinate and IMU sensor coordinate, we can describe the change amplitude curve between specified coordinate points (in this case, points of two arms and waist), by considering the weight of each kind of sensor, we can get a more accurate result to make comparison with database, and gaining a higher reliability on human motion recognition.

The equation of final degree of change is:

$$\Delta P = (\Delta P_v / P_v^0 \cdot \alpha + \Delta P_I / P_I^0 \cdot \beta)/2 \tag{2}$$

where ΔP_v is the change of motion from visual side, ΔP_I is the change of motion from IMU sensor side, P_v^0 and P_I^0 are the initial states of the current time segment, α and β are weight coefficients for visual and IMU sides.

The directional distance of each axes in each sensor can be calculated by Equation (1), and the degree of change in each sensor can be evaluated by Equation (2).

4 Experiment

This section makes experiment to validate the feasibility of our proposal. It includes the following aspects:

(1) Using depth camera to detect human skeleton and motion based on our visual recognition method, and record the 3D coordinates of depth maps;
(2) Using attached IMU sensors to detect human upper body skeleton and motion based on our applied motion recognition method, and record the coordinate differences of each sensor;
(3) Using our multi-type signal fusion correction method based on weight to generate coordinate differences from each frame and frequency. And output final accuracy.

In order to get the depth maps of subjects, an Intel Realsense D435i camera is settled, and shooting from one side of our experiment area. The picture of camera is shown in Figure 2, it can achieve a smooth video streaming with 848x480 resolution and 30 frames per second. Possible detecting range is from 0.3 m to 16 m.

In order to get the 3D motion data of subjects, 3 Witmotion IMU sensors are attached on subjects' two outer arms, and front waist. The arm sensor model is BWT901CL, the detectable parameters are acceleration, angle, velocity, magnitude, temperature. The waist sensor model is WTAHRS2, the detectable parameters include above and air pressure, height, which can make sure the surrounding environment of workers is stable and comfortable.

The experiment area is settled as Figure 3, depth camera is placed in front of the whole area, IMU sensors are attached on worker's outer arms and front waist, the worker will continue making different gestures in front of the depth camera.

Before the beginning of experiment, the database of actions to be tested which also is the control group will be prepared. In this experiment, several motions will be considered, including normal motions such as stand, sit, and sleep (lie down), other specific motions in construction site such as uplift, pick up heavy object, hold up heavy object, arms raised, regular cyclical movement and bend over will also be included.

The experiment process will be introduced as follows:

First, experiment subject (worker) will be attached with sensors, and stand in the proper position inside experiment area;

Next, experiment subject will make corresponding actions in order, there will be a break between each two motions;

Then, depth camera-based method will generate depth map, all pixels' coordinate information will be recorded, human skeleton will be generated based on image processing, human parts will be labeled and

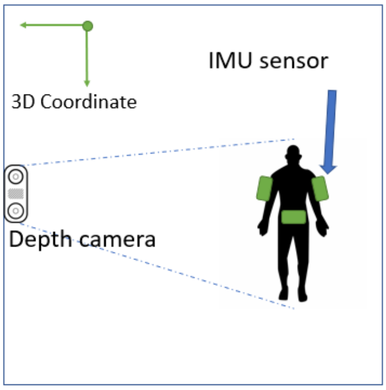

Figure 3. experiment area layout

corresponding to pixels, the coordinate amplitude of these parts will be recorded as well;

Next, IMUs-based method will collect data changes from 3 sensors, during the process, a low pass filter will be settled to eliminate redundant noise;

Then, the acceleration changes of each sensors will be used to calculate path changes by double integral;

Next, the changes of points of three human parts from visual side and IMUs side will be calculated separately to obtain degree of change within a certain period of time;

Then, degree of change from visual and IMUs side will be used to calculate the weight average, the result will be compared with database to find the best match;

Meanwhile, origin degree of change from visual and IMUs side will be compared with database separately;

Finally, the similarity from visual side, IMUs side and sensor fusion side will be compared, to prove if the sensor fusion method shows the best result.

As for the degree of change for IMUs, we set experimental steps as follows:

1. Calibration procedure: remove the output offset component of the acceleration sensor because of the presence of static acceleration (gravity). The method is to average the acceleration when there is no motion in accelerometer. (the more samples we collect, the more accurate the calibration result will be)
2. Low pass filtering: eliminate signal noise in accelerometers (both mechanical and electronic), to decrease the error while integrating the signal.
3. Mechanical filtering: when in a stationary state, small errors in acceleration will be treated as constant speeds, it indicates a continuous movement and unstable position, which will affect the actual motion detection. A mechanical filtering window will help to distinguish the small errors.
4. Positioning: the acceleration of each time period is known, we use double integral to obtain distance

information. The first integral gains speed and the second gains position.

In our simulation, we suppose 9 motions have their own perfect degree of change, includes vector changes from left arm, right arm and front waist:

$$M_n = \begin{bmatrix} \overrightarrow{L_n} \\ \overrightarrow{R_n} \\ \overrightarrow{W_n} \end{bmatrix}, n \in [1,9] \tag{3}$$

$$\overrightarrow{L_n} = \left(\Delta x_n^L, \Delta y_n^L, \Delta z_n^L\right) \tag{4}$$

$$\overrightarrow{R_n} = \left(\Delta x_n^R, \Delta y_n^R, \Delta z_n^R\right) \tag{5}$$

$$\overrightarrow{W_n} = \left(\Delta x_n^W, \Delta y_n^W, \Delta z_n^W\right) \tag{6}$$

where $\vec{L}, \vec{R}, \vec{W}$ represent the change of motion from left arm, right arm and front waist.

Generated data are divided into 2 groups, on visual side, the vector changes will be found by pixels and depth, on IMUs side, through the acceleration of three axes and time, Equation (1) will be used to obtain the distance in all directions, thereby obtaining the vector change. Then, assign a weighting coefficient to vision and IMU part through standard normal distribution, next, we will use Equation (2) to calculate the integrated vector change:

$$F = \begin{bmatrix} (L_v/L_v^0 \cdot \alpha + L_I/L_I^0 \cdot \beta)/2 \\ (R_v/R_v^0 \cdot \alpha + R_I/R_I^0 \cdot \beta)/2 \\ (W_v/W_v^0 \cdot \alpha + W_I/W_I^0 \cdot \beta)/2 \end{bmatrix} \tag{7}$$

where L_v is the left arm vector change on visual side, L_I is the left arm vector change on IMU side. F will be compared with M_1 to M_9 in Equation (3) to find out the best match.

100 pairs of sample data for each motion are generated based on our database by adding random interferences and white noise, to simulate deviations caused by the effects of real data collection. Weight coefficients obey standard normal distribution. By comparing our modified data with real motion data, the result of accuracy is shown on Figure 4. We can see that some similar motions such as standing and pick up heavy object, hold up heavy object and arm raised, sometimes are indistinguishable, this may become a more serious problem in real world sampling. Our simulation on 900 pairs of motion samples in total shows that averagely the accuracy of motion recognition by multi-type sensor fusion is about 97%, although the number of samples is not large, but this result shows that multi-type sensor fusion is possible to improve the accuracy of specific motion recognition in construction site condition. In real world experiment, due to other unexpected interferences and noises, the result may

real motion / detected motion	stand	sit	sleep	uplift	pick up heavy	hold up heavy	arm raised	regular cyclical	bend over	total
stand	95	3			2					
sit		96			1					
sleeping			100							
uplift				98			1			
pick up heavy	5	1			96					
hold up heavy						92	5			
arm raised				2	1	8	94			
regular cyclical								100		
bend over									100	
accuracy	95%	96%	100%	98%	96%	92%	94%	100%	100%	**97%**

Figure 4. simulation result.

change a bit, but due to the large number of samples, we expect the result to be as good as our simulation result.

5 Conclusion

In this paper, a new method is proposed for motion recognition and safety management of construction workers by using integrated sensor network, in order to effectively ensure the safety of workers in complex environment of construction site. We improve the motion recognition with depth maps, we also proposed several new motions that are usually shown in construction site and generate the database of each new motions' degree of change in relative 3D coordinate system. We proved that using multi-type sensor fusion to recognize human motion is possible, and our simulation shows that the accuracy is quite high compare to some related works.

We also noticed several problems during our research, such as distinction of similar motions. In the future, we will consider to add more special motions to expand recognition range, we will also improve the accuracy of skeleton interest point movement to decrease detection error from similar motions. IMUs-based motion recognition method is considered to be improved as well, meanwhile, a remote VR-based motion recognition system is included in our consideration.

References

[1] ACCESS, Comparison of Japan with other developed countries, Online: http://www.kasetsuanzen.or.jp/industrial_accident/compare.html, Accessed: 08/06/2020.

[2] JCOSHA, Occurrence of labor disaster in construction, Online: https://www.kensaibou.or.jp/safe_tech/statistics/occupational_accidents.html, Accessed: 08/06/2020.

[3] J. F. Yin, et al., Depth Maps Restoration for Human Using RealSense. IEEE Access, Vol. 7, pages 112544-112553, 2019.

[4] O. C. Ann, et al., Human activity recognition: A

review, IEEE International Conference on Control System, Computing and Engineering (ICCSCE 2014), pages 389–393, Batu Ferringhi, Malaysia, Nov. 2014.
[5] O. Dehzangi, et al., "IMU-based gait recognition using convolutional neural networks and multi-sensor fusion, Sensors, vol. 17, no. 12, page 2735, 2017.
[6] A. Cismas, et al., Crash detection using imu sensors, 21st International Conference on Control Systems and Computer Science (CSCS), pages 672–676, Bucharest, Romania, May 2017.
[7] S. J. Preece, et al., A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data, IEEE Transactions on Biomedical Engineering, vol. 56, no. 3, pages 871–879, March 2009.
[8] S. J. Preece, et al., Activity identification using body mounted sensors a review of classification techniques, Physiological Measurement, vol. 30, no. 4, page R1, 2009.
[9] N. Sarafianos, et al., 3d human pose estimation: A review of the literature and analysis of covariates. Computer Vision and Image Understanding, 152, pages 1–20, 2016.
[10] L. Ballan, et al., Motion capture of hands in action using discriminative salient points. Computer Vision–ECCV 2012, pages 640–653, 2012.
[11] Y. H. Huang, et al., Towards Accurate Markerless Human Shape and Pose Estimation over Time. International Conference on 3D Vision (3DV), pages 421–430, 2017.
[12] H. Rhodin, et al., A versatile scene model with differentiable visibility applied to generative pose estimation. In Proceedings of the IEEE International Conference on Computer Vision, pages 765–773,2015.
[13] A. Aguia, et al., MARCOnI-ConvNet-Based MARker-Less Motion Capture in Outdoor and Indoor Scenes. IEEE transactions on pattern analysis and machine intelligence 39, 3, pages 501–514, 2017.
[14] D. Eigen, et al., Predicting depth, surface normal and semantic labels with a common multi-scale convolutional architecture, In Proc. IEEE Int. Conf. Comput. Vis., pages 2650-2658, Dec. 2015.
[15] B. Li, et al., Depth and surface normal estimation from monocular images using regression on deep features and hierarchical CRFs, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2015, pages 1119-1127.
[16] F. Liu, et al., Deep convolutional neural fields for depth estimation from a single image, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pages 5162-5170, Jun. 2015.
[17] Z. Zhang, et al., Monocular object instance segmentation and depth ordering with CNNs, arXiv:1505.03159, Dec. 2015.
[18] X. Zhou, et al., Towards 3D human pose estimation in the wild: A weakly-supervised approach, In Proc. IEEE Int. Conf. Comput. Vis., pages 398-407, Oct. 2017.
[19] K. Lai, et al., A large-scale hierarchical multiview RGB-D object dataset, In Proc. IEEE Int. Conf. Robot. Automat., pages 1817-1824, May 2011.
[20] M. Camplani, et al., Efficient spatio-temporal hole filling strategy for Kinect depth maps, Proc. SPIE, vol. 8290, Art. no. 82900E, Jan. 2012.
[21] S. Matyunin, et al., Temporal filtering for depth maps generated by Kinect depth camera, In Proc. 3DTV Conf., True Vis.-Capture, Transmiss. Display 3D Video, May 2011, pages 1-4.
[22] A. Telea, An image inpainting technique based on the fast marching method, J. Graph. Tools, vol. 9, no. 1, pages 23-34, 2004.
[23] D. Miao, et al., Texture-assisted Kinect depth inpainting, In Proc. IEEE Int. Symp. Circuits Syst., pages 604-607, May 2012.
[24] D. Anguelov, et al., SCAPE: Shape completion and animation of people, ACM Trans. Graph., vol. 24, no. 3, pages 408-416, Jul. 2005.
[25] A. Weiss, et al., Home 3D body scans from noisy image and range data, in Proc. Int. Conf. Comput. Vis., pages 1951-1958, Nov. 2011.
[26] L. Liao, et al., Individual 3D model estimation for realtime human motion capture, in Proc. Int. Conf. Virtual Reality Vis., pages 235-240, Oct. 2017.
[27] D. Roetenberg, et al., Moven: Full 6dof human motion tracking using miniature inertial sensors. Xsen Technologies, December 2007.
[28] T. Marcard, et al., Sparse inertial poser: Automatic 3D human pose estimation from sparse IMUs. In Computer Graphics Forum, Vol. 36. Wiley Online Library, pages 349–360, 2017.
[29] M. Loper, et al., SMPL: A skinned multi-person linear model. ACM Transactions on Graphics (TOG) Vol.34, no.6, page 248, 2015.
[30] G. Varol, et al., Learning from synthetic humans, In Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pages 109-117, Jul. 2017.
[31] O. Dehzangi, et al., "IMU-Based Robust Human Activity Recognition using Feature Analysis, Extraction, and Reduction", ICPR 2018, pages 1402-1408, Beijing, China, 2018.