

<https://doi.org/10.17576/apjitm-2020-0901-04>
Asia-Pacific Journal of Information Technology and Multimedia
Jurnal Teknologi Maklumat dan Multimedia Asia-Pasifik
Vol. 9 No. 1, June 2020: 40 - 57
e-ISSN: 2289-2192

Received: 22 April 2019
Accepted: 23 August 2019
Published: 15 June 2020

COMPARATIVE STUDY OF 3D RECONSTRUCTION METHODS FROM 2D SEQUENTIAL IMAGES IN SPORTS

AZRULHIZAM SHAPI'I
SANAZ PICHAK
MOHD SYAZWAN BAHARUDDIN
HIROYUKI IIDA

ABSTRACT

The process of 3D reconstruction is a basic problem in Computer Vision. However, recent researches have been successfully addressed by motion capture systems with body worn markers and multiple cameras. To recover 3D reconstruction from fully-body human pose by single camera still remains a challenging problem. For instance, noisy background, variation in human appearance and self-occlusion were among these challenges. This thesis investigated methods of 3D reconstruction from monocular image sequences in dynamic activities such as sports. Six recent methods were selected based on they focused on recovery fully automated system for estimating 3D human pose for 2D joint location. These researches have been developed the algorithm that be able to solve ill-posed problem. Evaluation of the methods was divided in two sections. First, the theoretical and comparative study of each method was disclosed to identify the technique used, the problems that enquired and the results achieved in their approach. After that, the advantages and disadvantages of each method were listed. Also, several factors such as accuracy, self-occlusion and so on have been compared amongst these methods. In Second stage, based on the advantages found in the first stage of evaluation, three methods were chosen to be evaluated using specific data set. Initially, the codes of the three methods on PennAction dataset (tennis) were run and the performance of the methods in 3D reconstruction is showed. Then, the methods were tested on a mixed activities sequence from the CMU motion capture database. The novel of this study is evaluation of recent methods based on the accuracy of their performance on the specific dataset of tennis player. Also, we proposed a technique which combining specific advantages of each method to create a more efficient method for 3D reconstruction of 2D sequential images in the context of outdoor activities.

Keywords: 3D reconstruction sport pose, human pose, images sequence.

INTRODUCTION

Multimedia equipment can capture video or multi-photographs in real time in the course of a sport activity that can be replayed to an athlete player after the game to identify and rectify faults in technique. However, although this technique is flexible, the images shown provide only a single perspective (single camera view) which reduces considerably the ability to conduct an in-depth analysis (Thompson et al 2014). To address this issue, multiple cameras can be used for simultaneous capture of the player's performance, but this will incur high cost and be complicated. It will also require post processing and thus limit the time for motion capture. On the other hand, multiple challenges must be simplified in 3D reconstruction of human body area from sequential images.

In this article some considerations are taken into account of the different methods analyzed in order to determine the most suitable method to be applied for the sport of tennis. First, "realistic human body" has been targeted due to the complexity in modeling based on variations in individual body shape and different clothes. Second, the accurate recognition of self-occlusion where some limbs block other body parts in the images and obstruct the vision

when stationary camera is studied. Third, finding proper image descriptors can be more helpful in resolving many pose ambiguities and usually require trial and evaluation procedure to determine most competitive representations. Finally, special attention was given to the inclusion of real-world conditions such as cluttered background, uncontrolled scenes, noisy data, speed of moving person in sequential frame (Saima et al. 2018). Therefore, this research identified the best method which improves the ill-posed problem and to the handle outdoor conditions in order to be implemented in the tennis court environment. This paper has three objectives which are as follows; first, we evaluate different methods for 3D reconstruction of the human body from a sequence of monocular images to determine which one performs efficiently under occlusions, noise on the real world data. Second, we compare between the developed and implemented 3D reconstruction methods, with the identification of the advantages and drawbacks of each. Third, we propose a new technique combining the advantages of different methods studied that is more accurate in a particular application (tennis sport) with fixed and ordinary camera.

LITERATURE REVIEW

Considerable research has addressed the challenge of human motion capture from imagery such as Gotardo & Martinez (2011), Ramakrishna et al. (2012) and Wandt et al. (2016) allow reconstructing 3D human motion using feature tracks in monocular image sequences and combining random camera motion depending on prior trained base poses. Also, they focuses on any motion; periodic and non-periodic.

The review of the methods was conducted following the method proposed by Ramakrishna et al. (2012), Atul (2014), Akhter et al. (2015), Xiaowei et al. (2016), Wandt et al.(2016) and Du et al.(2016). These six of the most recent 3D reconstruction algorithms were selected for the analysis based on the performance and result described in their research. The theoretical approach of all methods is discussed and the detail performance of the mathematical model was identified.

Ramakrishna et al. (2012) offered a model that was not activity-dependent to retrieve the 3D configuration of a human figure from 2D locations of anatomical points in a single image, leveraging a large motion capture corpus as a substitute for visual memory.

Atul (2014) developed three principled approaches to enhance particle filtering by integrating bottom-up information either as proposal density for obtaining more diverse particles or as complementary cues to improve likelihood computation during the correction step. In addition, he also demonstrated that a feedback mechanism from top-down modeling can further adapt and enhance the bottom-up predictors to improve tracking performance.

Akhter et al. (2015) modeled how joint-limits differ with pose for getting valid poses. They collected a motion capture dataset that explored a multiplicity of human poses and developed a pose-dependent model of joint limits that forms their prior.

Xiaowei et al. (2016) proposed the integration of a sparsity-driven 3D geometric prior and temporal smoothness when the image locations of the human joints are provided and when they are unknown, and this was extended by programming the image locations of the joints as latent variables by considering several ambiguities in 2D joint locations.

The approach suggested by Wandt et al. (2016) aims to address the issue of predicting non-rigid human 3D shape and motion from image sequences captured by non-calibrated cameras. They factorized 2D observations in camera parameters, base poses and mixing coefficients, in the same way as other state-of-the-art solutions. The novelty of this method compared with existing methods is that it can handle arbitrary camera motion without the need to use predefined skeleton or anthropometric constraints whereas other methods require adequate camera motion during the sequence to obtain a proper 3D reconstruction.

Du et al. (2016) proposed in their method the goal to make the 3D motion reconstruction more accurate, and so more built-in knowledge was added, such as height-map, was introduced into the algorithmic scheme of reconstructing the 3D pose/motion in a single-view calibrated camera.

Finally, our approach was a comparative study of 3D reconstruction methods of human body from 2D image sequence of tennis player. We focused on the evaluation of different methods that studied on sports poses by analyzing several factors such as accuracy of human pose estimation, self-occlusion, and noisy background that are still not fully resolved. So, we run code of their algorithm of these methods in MATLAB on PennAction dataset to get 3D reconstruction result. After collecting all the results and comparing them together, we proposed new technique which is combining of some approaches of three methods Xiaowei method, Wandt method and Du method. The novel method proposed be able to improve 2D joint location and occlusion that can recreate 2D images into 3D images with realistic results, minimum requirements and effective results.

METHODOLOGY

The methodology used in this research consisted of four phases described in Figure 1. Phase one, consisted in the analysis of multiple methods recently published for 3D reconstruction in order to identify six methods that showed to be the most relevant for the purpose of this research. Phase two, consisted in the comparison of the experimental result of each method presented by the authors based on several factors such as projection, camera, realistic reconstruction, self-occlusion, accuracy, noisy background and process speed in order to shorten the list to three highlighted methods. In phase three, the evaluation of these three selected methods was studied using a specific sequential images of tennis player data set and the results were compared. Finally, phase 4 consisted in the proposal of a new improved method for 3D reconstruction from 2D sequential images that combine the robustness of each method evaluated.

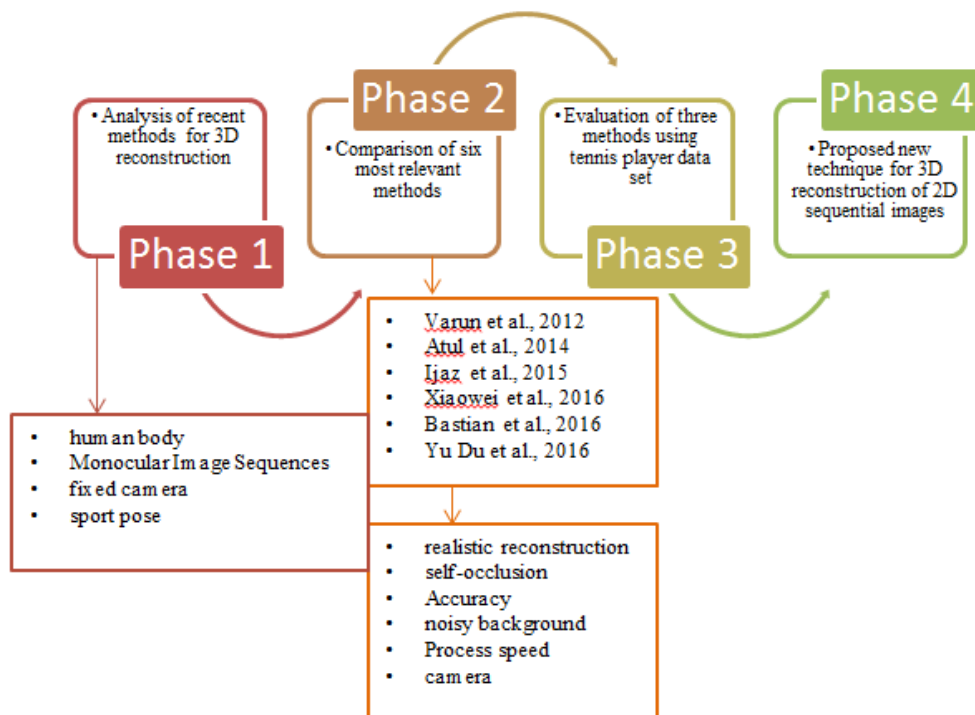


FIGURE 1. Methodology Use For Comparative Study of 3D Reconstruction

Figure 2 Displayed the pathway selected to evaluate the performance of the three selected methods. The first step for the analysis was conducted by analysing the mathematics described for each method. Following this, the code was diggited using MATLAB and the performance of each method was assessed using the specific dataset proposed by each author to verify that the codes working without error. However, when the code was not provided additional work was required and the mathematical analysis of the code was used to program the method as described by the author. Specific factors for these methods were evaluated on our particular data set to evaluate their performance and to compare their accuracy in 3D reconstruction (Li et al. 2013) . Output of running codes was compared using tennis player’s dataset (PennAction). Finally, these methods were evaluated on CMU dataset to understand their performance in 3D reconstruction error, accuracy percentage and to compare the results.

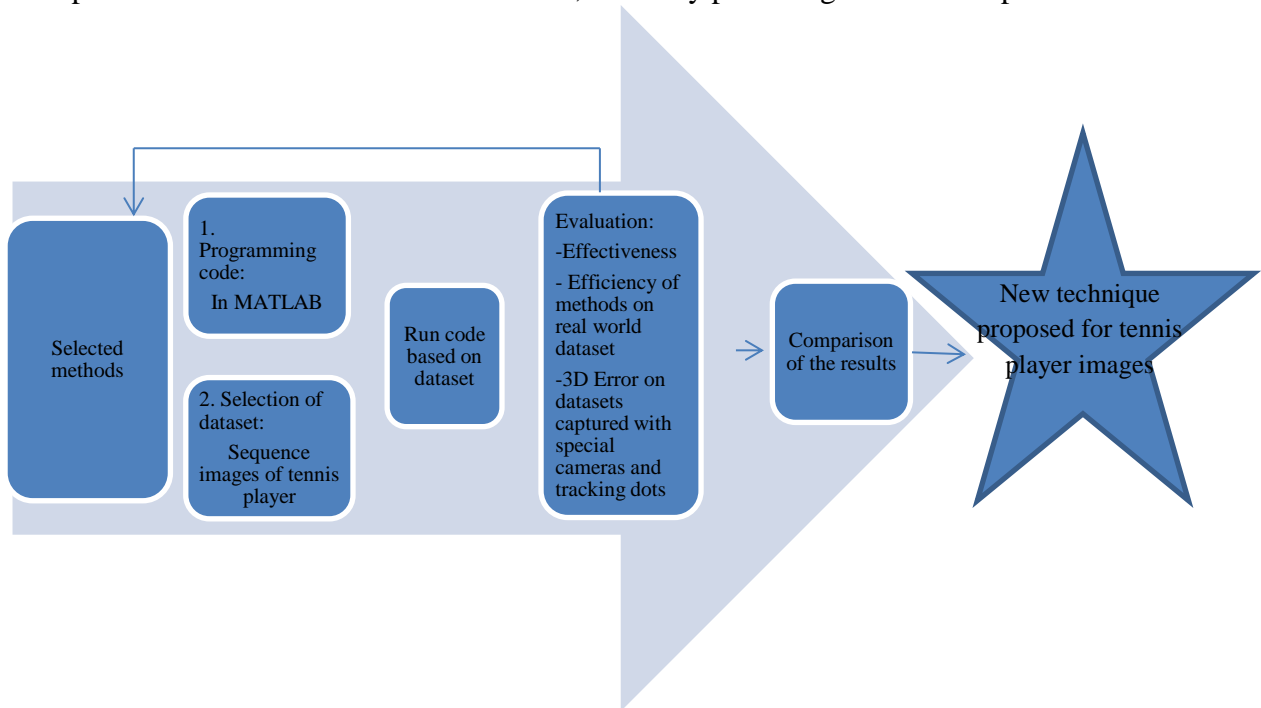


FIGURE 2. Process Used For The Evaluation of Three Selected Methods


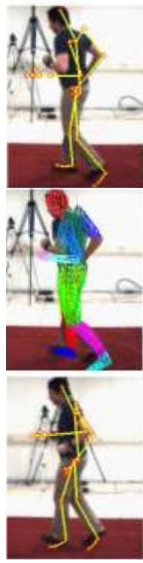
The final stage of this study consisted in the compilation of the advantage found in the evaluated methods. Specific advantages were integrated in the core method (i.e. The method that show the best performance on proposed data set) to overcome disadvantages found and to improve the efficacy for 3D reconstruction. The final method proposed include the highlights and provided a novel approach for 3D reconstruction from 2D sequential images in tennis sport.

RESULTS

The review of each selected method is presented in Table 1. The advantages and disadvantage of each one of their techniques are described.

TABLE 1. Summary Of Advantages And Disadvantages of Each Algorithm.

| Authors | Specification of application | Pros | Cons | Illustrations |
|---------|------------------------------|------|------|---------------|
|---------|------------------------------|------|------|---------------|

| | | | | |
|--------------------|---|--|--|---|
| Ramakrishna method | <p>1- Human pose recovery based on sparse representation in an over-complete dictionary</p> <p>2- enforces a mandatory criterion on the sum of squared limb lengths</p> <p>3-enforces the quantum of eight selected limbs for constancy.</p> <p>4- utilizes a model with 23 landmarks of the human anatomy.</p> <p>5- MP algorithm to estimate the sparse demonstration of 3D pose and the relative camera from only 2D image</p> | <p>1-solves anthropometric regularity</p> <p>2-robustness to missing data</p> <p>3-Joint Sensitivity to noise</p> <p>4- describes an expansive range of actions by a statistical model of human pose</p> <p>5- solves the pose and camera by reducing the image reprojection error.</p> <p>6- low RMS reconstruction error</p> <p>7- provides accurate results using a single image and no requirement for annotation to resolve ambiguities</p> <p>8-accuracy in the camera estimation</p> <p>9- applied to frames of monocular video streams</p> <p>10-able to recover the pose from non-standard viewpoints</p> <p>11- good generalization to a large range of poses and viewpoints</p> | <p>1-Limb proportions are different between various individuals.</p> <p>2- Does not support occlusion handling.</p> <p>3-Cannot recover the correct pose with good perspective effects when the mean pose is not a reasonable initialization</p> |  |
| Atul method | <p>1-fully automated 3D human pose and shape analysis of the human targets in videos, recognizing their activities and characterizing their behavior</p> <p>2- combines Top-down and Bottom-up methods</p> <p>3-uses advanced Particle filtering (PF) algorithms.</p> <p>4- uses framework of non-parametric density propagation system based on particle filtering</p> | <p>1- high efficacy in strong ambiguities</p> <p>2- overcomes limitations of particle filtering by improving the proposal density modeling and the likelihood computation</p> <p>3-improves tracking</p> <p>4-solves non-rigid deformable surface reconstruction</p> <p>5- articulates body pose recovery in static images</p> | <p>1-self and Partial occlusion in unseen scenarios</p> <p>2-Optimization problems</p> <p>3- uses fixed bone lengths priors</p> |  |

Akhter method

1-a physically-motivated prior allows anthropometrically valid poses and restricts on invalid poses
2- prior is combined with a selected sparse representation of poses from an over-complete dictionary
3- formulates prior over two endpoints of each 3D bone location.
4- uses Orthogonal Matching Pursuit (OMP)

1- combines a sparsity-driven 3D geometric prior and a 3D temporal smoothness prior
2- uses a deep convolutional neural network (CNN) architecture to detect body parts
3- an Expectation-Maximization (EM) framework to retrieve a sparse model of 3D human pose sequence
4- casting the 2D joint locations as latent variables

Xiaowei method

1-A periodic model to mix coefficients for periodic and quasi-periodic motions
2- a regularization term based on temporal bone length constancy prior for non-periodic motion
3-based on a-priorly trained base poses

Wandt method

1-good generalization while avoiding invalid 3D poses
2-pose parameterization is simple and accurate
3-improves pose estimation by a grouping of body parts (extended-torso)
4-defines a kinematic skeleton tree structure to apply joint-angle limits
5-avoids non-rigid self-calibration by selecting linear coefficients from a cosine function

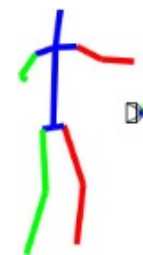
1- highly effective against detector error, occlusion, and ambiguity
2- no requirement for synchronized 2D-3D data
3- handles the 2D estimation uncertainty in a statistical framework
4-good accuracy in-the-wild videos
5- improves the initialization results
6- improves 2D joint localization
7- using a single camera





1-estimates non-rigid human body pose captured by an uncalibrated camera
2-solves an unstable 3D motion reconstruction
3-accurate algorithm for estimating periodic motion
4- handles arbitrary camera motion
5-the stability of method

1-depth ambiguities at several joints
2- incorrect estimation of the camera matrix
3-algorithm is sensitive against Gaussian noise
4-needs multi camera setups

1- cannot handle multiple subjects
2- assumes manually labelled 2D joint locations

1-restrictive assumptions on the 3D configurations possible
2- ambiguous camera placement and 3D shape deformation



| | | | | |
|-----------|---|---|---|---|
| | 4-model 3D pose as a linear combination of base poses | 6-handles noise and occlusions in real world data 7- does not use a predefined skeleton or anthropometric constraints | | |
| Du method | 1- RGB-based 2D joint detection algorithm 2- a dual-stream Deep Convolution Network (ConvNet) to detect 2D landmarks of human joints | 1- improves skeleton-based human 3D pose estimation from the inaccurate localization of 2D joints by temporal constraints on the camera 2-lower reconstruction error | 1-discriminative parts are missed in pose estimation 2-fails in complex human motion |     |
| | 3-algorithm utilizes the height-map as a type of built-in prior knowledge to capture 3D articulated skeleton motion under a single-view calibrated camera | | | |

After identifying the weaknesses and strengths of each algorithm, their experimental results of different parameters are presented in Table 2. These results are based on publisher's own assessment of their selective database. For example, Ramakrishna et al (2012), Xiaowei et al (2016) and Bastian et al (2016) focused their analysis in noisy background. On the other hand, Atul (2014) et al, Akhter et al (2015) and Du et al (2016), focused their results on self-occlusion. Moreover, some of the described methods present the evaluation on a single image whereas others evaluate on sequential images.

Three methods were found suitable for the proposed application of this study as described below. Among these methods, two methods (i.e. Xiaowei and Bastian) showed successful outcomes in noisy background, self-occlusion and realistic reconstruction, which makes them ideal to be further evaluated for 3D reconstruction. Furthermore, from the other methods compared, Yu du method was selected due to its outstanding results the parameter required for the analysis of the dataset chose. No significant advantages were found in the other methods analyzed and were discarded as they presented lack of noise background reduction, realistic reconstruction and both. Finally, these three selected methods showed to be the most suitable methods to be analyzed using their result in database of tennis's player.

TABLE 2. Comparison of Methods for 3D Reconstruction.

| Parameters | Reference | | | | | |
|------------|---------------------------------|---------------------------------|-----------------------------------|--------------------------|-------------------------|------------|
| | Ramakrishna Method | Atul Method | Akhter Method | Xiaowei Method | Wandt Method | Du Method |
| Algorithm | Projected Matching Pursuit (MP) | Advanced particle filtering(PF) | Orthogonal Matching Pursuit (OMP) | Expectation-Maximization | Periodic - non Periodic | Height-map |

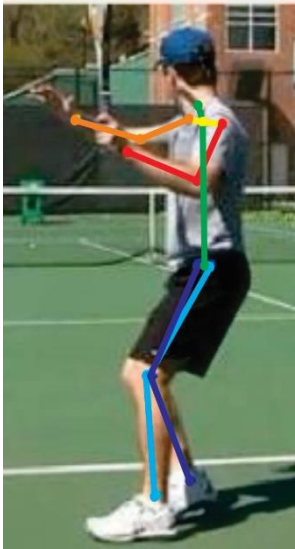
| | | | | | | |
|---|--------------------|-------------------|--------------------|-----------------------|-----------------|-------------------|
| Projection | Single image | Sequence images | Single image | Sequence images | Sequence images | Sequence images |
| Camera Code of algorithm | arbitrary restrict | Fixed Open source | arbitrary restrict | Arbitrary Open source | fixed restrict | fixed Open source |
| Noisy background | focus | No focus | No focus | focus | focus | No focus |
| Self-occlusion | No focus | Focus | No focus | focus | focus | focus |
| Realistic reconstruction (Average 3D error/ cm) | Not given | Not given | 121.56 | 113.01 | 187.1 | 118.69 |
| Average 2D joints localization accuracy (pixel) | 94.5 | 27.65 | Not given | 10.85 | 8.43 | 95.8 |

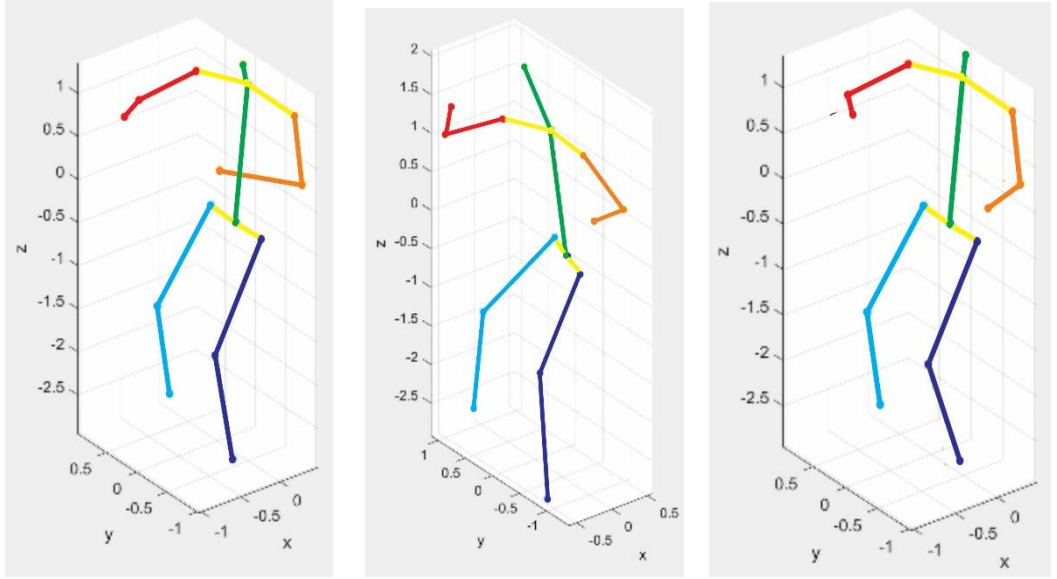
EVALUTION OF METHODS ON PENNACTION

This section demonstrates the application of the selected approaches for pose estimation with in-the-wild images sequence. Results are presented utilizing action from the PennAction dataset. The “tennis forehand” was selected for evaluation due to it is not simple pose. It also has some challenging such as the large pose variability, self-occlusion, and image blur because of fast motion. We selected six frames (2,8,14,20,25,30) from 31 images sequence of dataset that we were able to evaluate main factors.

Table 3 to 8 illustrated the 3D results of each method on frames.

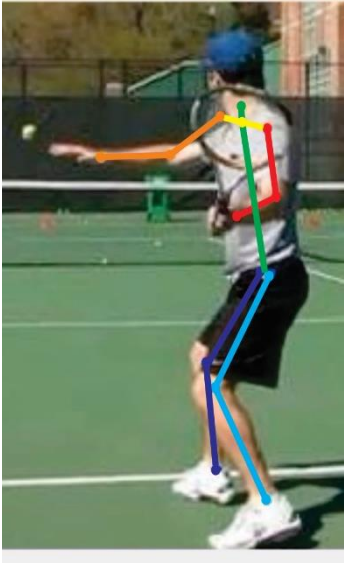
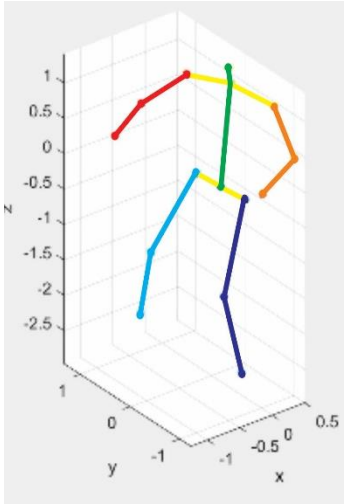
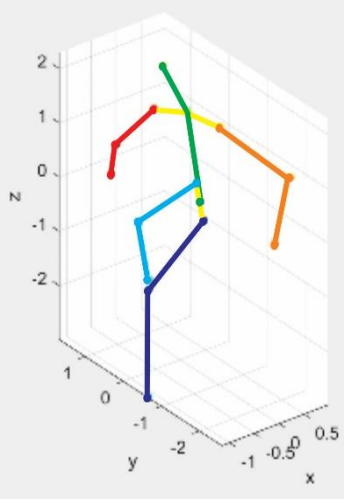
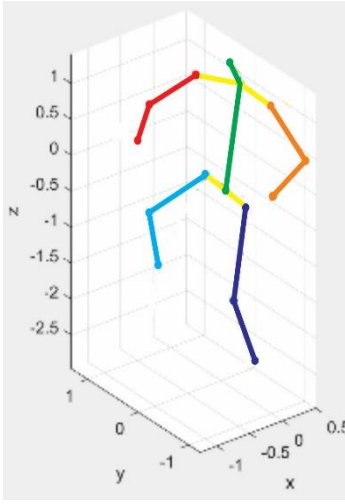
TABLE 3. 3D Results of Frame #2

| Input | Frame #2 | Xiaowei | Wandt | Du |
|-------|---|---------|-------|----|
| |  | | | |



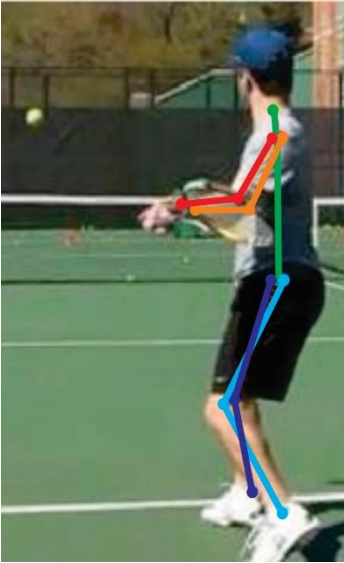
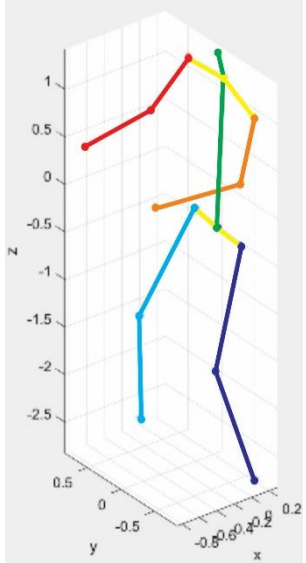
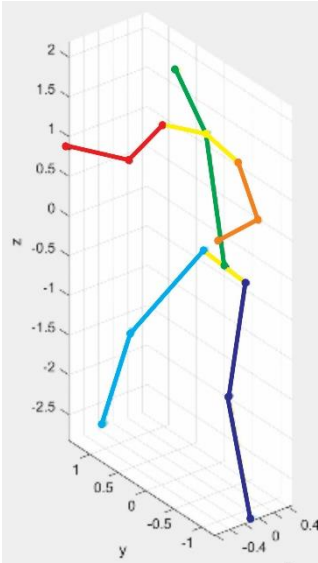
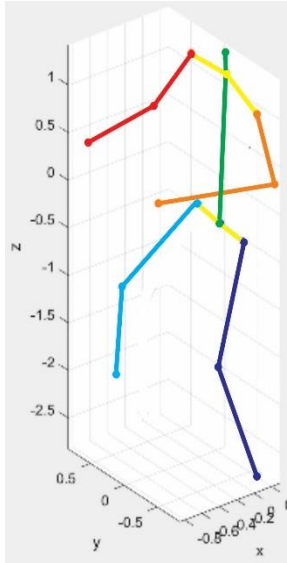
Analysis of frame #2 is shown that Wandt method didn't have correct result in left arm (red color) because of racquet as noisy environment.

TABLE 4. 3D Results of Frame #8

| Input | Frame #8 | | |
|---|--|---|--|
|  | Xiaowei | Wandt | Du |
| |  |  |  |

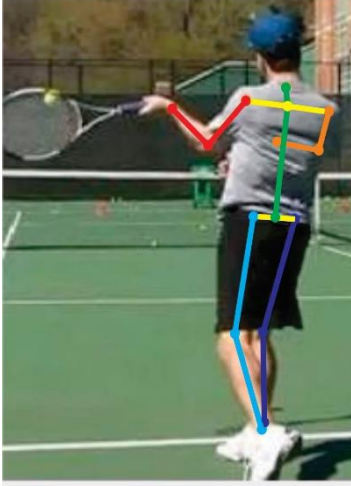
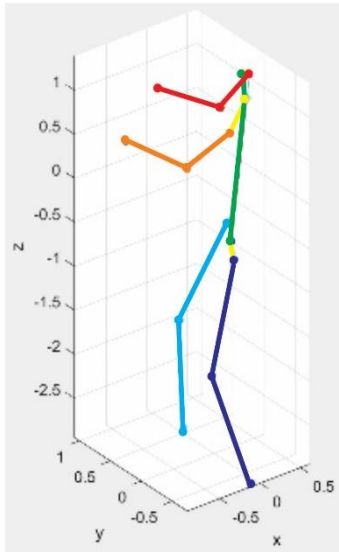
Analysis of frame #8 is shown that Du method didn't have accurate result in right leg (violet color) because of right leg occluded by left leg that it causes problem of self-occlusion.

TABLE 5. 3D Results of Frame #14

| Input | Frame #14 | | |
|---|--|---|--|
|  | | | |
| | Xiaowei | Wandt | Du |
| |  |  |  |

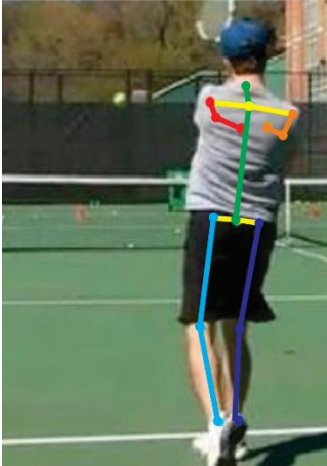
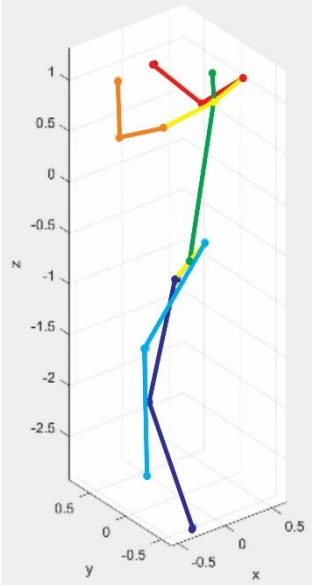
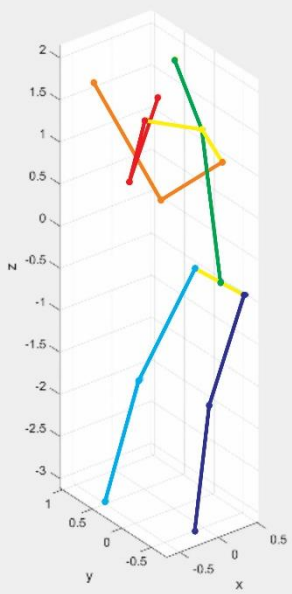
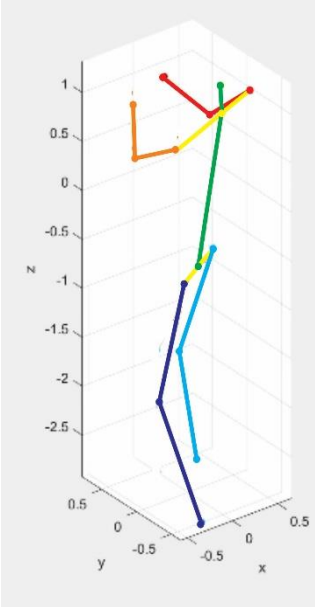
Analysis of frame #14 is shown that both methods Wandt and Du didn't have good result in both arms because of arms of player are behind racquet. It causes occlusion.

TABLE 6. 3D Results of Frame #20

| Input | Frame #20 | | |
|--|-----------|-------|----|
|  | Xiaowei | Wandt | Du |
|  | | | |

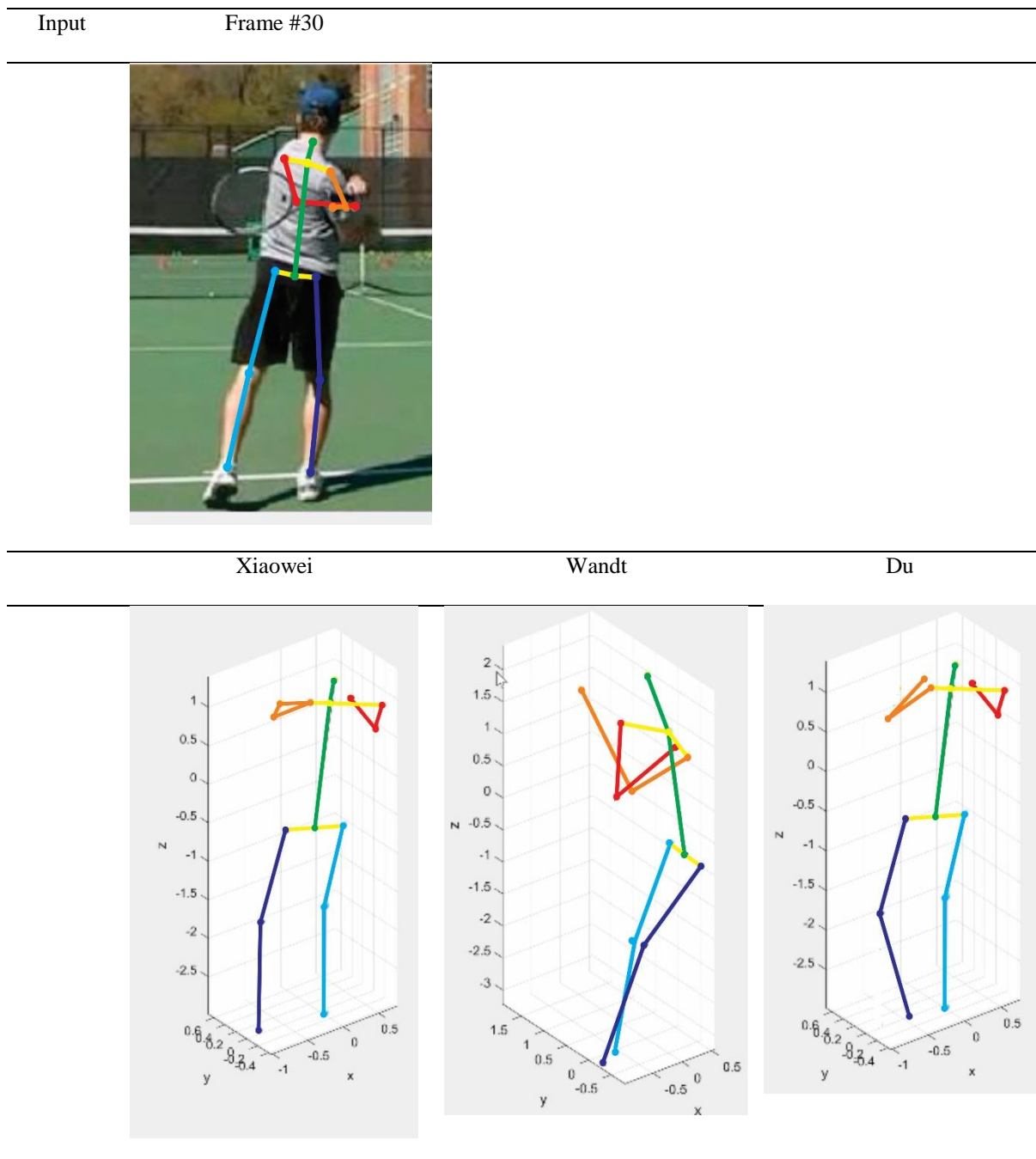
Analysis of frame #20 is shown that Xiaowei method had the best result in this specific angle with less missing data.

TABLE 7. 3D Results of Frame #25

| Input | Frame #25 | | |
|---|--|--|--|
|  | | | |
| | Xiaowei | Wandt | Du |
| |  |  |  |

Analysis of frame #25 is shown that Wandt method had poor result in arms and shoulder (yellow color) but had better result in the parts of legs compare with Du method.

TABLE 8. 3D Results of Frame #30



Analysis of frame #30 is shown that Wandt method is sensitive to angle of image and couldn't reconstruct with accuracy.

Table 9 shows a summary of results on tennis player dataset. The conclusion of these results is shown that the method proposed by Xiaowei et al. (2016) and Du et al. (2016) had several similarity in 3D result. But Xiaowei algorithm is more robust to noise and also able to handle occlusions and reconstruct the occluded body parts correctly. Although, Bastian's method revealed better performance to reconstruct in the part of legs.

Table 9. A comparison of PCP scores on PennAction dataset

| Methods | Upper Arms | Lower Arms | Upper legs | Lower legs | Average |
|---------|------------|------------|------------|------------|---------|
| Xiaowei | 0.93 | 0.71 | 0.96 | 0.84 | 0.86 |
| Du | 0.92 | 0.68 | 0.94 | 0.82 | 0.84 |
| Wandt | 0.90 | 0.61 | 0.98 | 0.85 | 0.83 |

RUNNING TIME

Table 10 showed the processing time of the three methods evaluated. It is clear that the method proposed by Xiaowei is fastest method to process 3D reconstruction according to the specifications mentioned in the computing section. Algorithms usually converge in 20 iterations with average CPU time below 150s for a sequence of 31 frames.

TABLE 10. The Processing Time of Each Algorithm

| | processing time (seconds) per frame |
|----------------|-------------------------------------|
| Xiaowei method | 3 |
| Wandt method | 6 |
| Du method | 5 |

EVALUATION OF METHODS ON CMU

These methods were evaluated by testing them on a sequence of mixed activities from the CMU motion capture database. Care was taken to make sure that the motion capture frames were not those utilized in the training of the shape bases. It could be seen that the reconstruction results of the jumping sequences were inferior in comparison with the other sequences. This was because the difference between jumping motions of various individuals was much larger than between running motions. As such, a new, untrained jumping motion was insufficiently explained by the base poses, whereas each new running pattern was the same as those in the training data. Second, the evaluation of 3D motion recovery was carried out with the ground-truth 2D joint locations.

The 3D reconstruction errors in millimeters are reported in Table 11. The standard evaluation per joint error (mm) in 3D was computed between the reconstructed pose and the ground truth in the camera frame and their root locations aligned.

TABLE 11. 3D reconstruction error in mm

| Method | run (mm) | Jump (mm) |
|---------|-------------|--------------|
| Wandt | 28.05 | 31.13 |
| Du | 58.3 | 64.4 |
| Xiaowei | 20.99 | 47.57 |

This table is shown 3D reconstructions of Xiaowei are highly realistic, which was shown by the 3D error.

DISCUSSION

The result from the evaluation conducted indicated that the method proposed by Xiaowei et al. (2016) is highly recommended for 3D reconstruction of tennis player images. In this method, 2D joint heat maps capturing positional uncertainty are generated with a deep fully CNN. These heat maps are combined with a sparse model of 3D human pose within an Expectation-

Maximization framework realized the 3D parameter estimation over the entire sequence. However, this method provided a solution for most of the challenges in 3D reconstruction such as large pose variability, self-occlusion, and image blur caused by fast motion. But, it needs manually labeled for 2D joint location that reduces percentage of accuracy. To improve this issue, we proposed 3D human pose estimation framework that presented by Wandt et al. (2016) method. It consists a synthesis between discriminative image-based and 3D reconstruction. It treated 2D joint locations as latent variables whose uncertainty distributions are given by a deep fully convolutional neural network. The unknown 3D poses are modeled by a sparse representation and the 3D parameter estimates are realized via an Expectation-Maximization algorithm, where it is shown that the 2D joint location uncertainties can be conveniently marginalized out during inference. Further, to improve robustness of method against occlusion and reconstruction ambiguity, 3D temporal smoothness prior is imposed on the 3D pose and viewpoint parameters which Du et al. (2016) method considered. Therefore, the usage of the method proposed by Xiaowei et al. (2016) as core-base and the integration of the advantages described for the method by Wandt et al. (2016) and Du et al. (2016) might provide an effective method for 3D reconstruction of images sequence on specific dataset. The novel method proposed might be able to improve 2D joint location and occlusion that can recreate 2D images into 3D images with realistic results, minimum requirements and effective results.

CONCLUSION

This paper is a comparative study of 3D reconstruction methods of human body from 2D image sequence of tennis player. Among all the sports, we chose tennis sport because this exercise developed at very high-speed and requires the development of technical skills. Also, this sport presents a challenge for the 3D reconstruction due to factors such as self-occlusion and occlusion that occurs during the development of the game. Many technologies tried to help raise the level of athletic techniques; as well as reduce arbitration errors and physical damage but, they faced multiple problems such as high cost, time consuming and heavy equipment (Norshaliza, 2016). We believe that the simulation of tennis players' movement taken by arbitrary camera through 3D reconstruction of sequential images reduces might be economical viable and simplify the time when compared with traditional technologies. Moreover, this method can help the players and coaches to significantly improve skills. On the other hand, the increasing demand of 3D reconstruction especially for the human body can provide multiple additional applications such as movies, gaming and medical purpose. The achievement of this research can also help other industries as well. Specially, generating 3D poses from a sequence of images is much cheaper than marker-base technologies.

Modeling of 3D human body from image sequences is a challenging problem and has been a research topic for many years (Ashraf et al. 2014). Important theoretical and algorithmic results were achieved that allow to extract even complex poses of human body form. Research in the area of human pose has been approached from many different issues in an attempt to implement a robust, accurate and automatic fully-body system.

In this paper we focused on the evaluation of different methods that studied on sports poses by analyzing several factors that are still not fully resolved in this area. For instance, realistic scenes background clutter, variation in human appearance, and self-occlusion are challenges that require in deep investigation. Also, we identified the most suitable method which improves the ill-posed problem and can handle outdoor conditions in order to be implemented in the tennis court environment with high speed process (Shingade and Ghotkar, 2014). To reach this goal, there are two step of evaluation. First of all, we have chosen six recent methods based on their focus on several features such as image sequence, camera, sport

poses in real world and so on. These methods have improved several challenges in old methods and some recommendations for future work. Advantages and disadvantages of methods of Ramakrishna et al (2012), Atul (2014), I Akhter et al (2015), Xiaowei et al. (2016), Wandt et al. (2016) and Du et al. (2016) were discussed and compare theoretically. Some factors such as accuracy of human pose estimation, self-occlusion, and noisy background were analyzed in their experimental results.

In the next step of evaluation, three top methods are selected for further and deeper analyze. We run code of their algorithm in MATLAB on PennAction dataset to get 3D reconstruction result. The codes of Xiaowei et al. (2016) and Du et al. (2016) obtained from the Internet. We were implemented code of Wandt et al. (2016) method by ourselves. To get the final and definitive results, we also tested these methods on database CMU MoCap. After that, it was decided that among them, the methods proposed by Xiaowei might be the most suitable method to be implemented for the 3D reconstruction applied to tennis. This method proved to be faster than the other method evaluated and produced outstanding results in terms of accuracy. Subsequently, the method proposed by the Wandt showed to provide better accuracy when dealing with self-occlusions. Finally, the method proposed by Du showed the lowest accuracy and poor performance when occlusion was involved.

Eventually, we proposed new technique which is combining of some approaches of three methods Xiaowei method, Wandt method and Du method. It proposed a 3D human pose estimation framework from a monocular image that consists of a novel synthesis between a deep learning-based 2D part regressor, a sparsity-driven 3D reconstruction approach of Wandt method, and a 3D temporal smoothness prior in Du method. This joint consideration combines the discriminative power of state-of-the-art 2D part detectors, the expressiveness of 3D pose models, and regularization by way of aggregating information over time. So, it can go directly from 2D appearance to 3D geometry. Proposed method can improve 2D joint locations for tennis player poses in outdoor condition from sequence images taken by arbitrary camera.

REFERENCES

- Akhter, I. and Black, M. J. 2015. Pose-Conditioned Joint Angle Limits For 3D Human Pose Reconstruction. *IEEE Conference On Computer Vision And Pattern Recognition (CVPR)*, pp. 1446-1455
- Ashraf, Y. A., Venkat, I. and Belaton, B. 2014. Reconstruction of 3d Faces By Shape Estimation And Texture Interpolation. *Asia-Pacific Journal of Information Technology And Multimedia*, 3(1):pp. 15 – 21.
- Atul, K. 2014. Coupling Top-down and Bottom-up Methods for 3D Human Pose and Shape Estimation from Monocular Image Sequences. *Pattern Recognition* :1410-0117.
- Du, Y., Wong, Y., Liu, Y., Han, F., Gui, Y., Wang, Z., Kankanhalli, M. & Geng, W. 2016. Marker-less 3D human motion capture with monocular image sequence and height-maps. *In European Conference on Computer Vision*. pp. 20-36.
- Gotardo, P.F. & Martinez, A.M. 2011. Non-rigid structure from motion with complementary rank-3 spaces. *In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*: pp. 3065-3072.
- Norshaliza, K. 2016. Active Contour Model Using Fractional Sinc Wave Function For Medical Image Segmentation. *Asia-Pacific Journal of Information Technology and Multimedia*, 5(2): 47 – 61.
- Ramakrishna, V., Kanade, T. & Sheikh, Y. 2012. Reconstructing 3d human pose from 2d image landmarks. *In European Conference on Computer Vision*. pp. 573-586.
- Saima, A. L., Rosziati, I., Nik Shahidah, A. M. T., Norhalina, S. and Suhaila, S. 2018. Thresholding And Quantization Algorithms For Image Compression Techniques: A Review. *Asia-Pacific Journal of Information Technology and Multimedia*, 7(1) : 83 – 89.

- Shingade, A. & Ghotkar, A. 2014. Animation of 3D human model using markerless motion capture applied to sports. *International Journal of Computer Graphics & Animation (IJCGA)* 4(1): 27-39.
- Wandt, B., Ackermann, H. & Rosenhahn, B. 2016. 3D Reconstruction of Human Motion from Monocular Image Sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(8):pp. 1505-1516.
- Xiaowei, Z., X., Zhu, M., Leonardos, S., Derpanis, K.G. and Daniilidis, K. 2016. Sparseness meets deepness: 3D human pose estimation from monocular video. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 4966-4975.

Azrulhizam Shapi'i

Mohd Syazwan Baharuddin

Fakulti Teknologi & Sains Maklumat

Universiti Kebangsaan Malaysia

azrulhizam@ukm.edu.my, syazwan@ukm.edu.my

Sanaz Pichak

Faculty of Technical and Engineering

Islamic Azad University, Tehran, Iran

sanaz.pichak@yahoo.com

Hiroyuki Iida

Japan Advance Institute of Science and Technology

1 Chome-1 Asahidai, Nomi, Ishikawa 923-1211, Japan

iida@jaist.ac.jp