2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 25-29, 2020, Las Vegas, NV, USA (Virtual)

Whole-Game Motion Capturing of Team Sports: System Architecture and Integrated Calibration

Yosuke Ikegami[1], Milutin Nikolić[1,2], Ayaka Yamada[1], Lei Zhang[1,3], Natsu Ooke[1] and Yoshihiko Nakamura[1]

***Abstract*— This paper discusses the application of video motion capturing technology (VMocap) to a competitive team sports game. The setting introduces a specific set of constraints: large scale markerless motion capturing, big recording volume, transmitting and processing gigabytes of data, operation without interfering with players or distracting spectators and staff, etc... In this paper, we present how we tackled and successfully solved all of these constraints. That enabled us to analyze the sportsmen without any intrusions, while giving their peak performance, hence opening a new field for Mocap application. International volleyball game was recorded in full length with the described system. During the course of the event, we compressed 54TB of raw image data real-time, capturing 6 hours of high framerate video per camera, without disturbing any of the game operations. Using the data, we were able to reconstruct the motion, muscle activity and behavior of the athletes present on the court.**

I. INTRODUCTION

Studies on computational algorithms of complex mechanical systems in robotics have lead the advances of not only motion control of robotics systems, but also motion analysis of human movements. Inverse kinematics and inverse dynamics in robotics are now key technologies in biomechanics. Humanoid robotics also extended scientific knowledge of biped locomotion. The technologies of segmentation and classification of motion data of human developed in humanoid robotics find their application in data science of human motion. Agile technology developed in robotics and related fields such as deep neural networks, computer vision, high-speed data acquisition, network communication, and realtime computation will find many applications to enhance the human life and society. This paper presents the development of a prototype of video motion capture system that can capture the whole game of competitive team sports matches in a large space. The system can be used for the 3D reconstruction of athletes' motion and its biomechanical and behavioral analysis.

Since it became commercially available, the motion capture technology (Mocap) found numerous applications in industry and research. Most of applications exploit the capability to capture the human motion. Namely, the technology is widely used in rehabilitation, sports science, animation/movie/game industry, human motion analysis/recognition [1], [2], [3] and human behavior modeling [4], [5]. Mocap is also employed in non-human related fields where external motion tracking is required [6], [7].

The most common Mocap technologies are passive-optical, active-optical and IMU-based motion capture. Optical motion capturing systems are using multi-camera setup and triangulation to determine the positions of markers. Markers can be be passive (reflective) that are light up by infrared light [8], [9], or active, usually stroboscopic LED lights [10]. These kinds of systems usually need around 40 markers to capture a single person, thus the measurements are bound to take place in laboratories. IMU-based Mocap technology avoids this problem by reconstructing the human skeleton by using measurements from IMU sensors placed on the subjects body [11]. Hence, the recording is not constrained to the laboratory setting but can be performed in the everyday environment of humans. These systems are usually less accurate and magnetometer readings can be disturbed by ferrite objects in the environment, causing additional inaccuracies.

Above mentioned methods need an array of devices to be attached to the human body which cause two major drawbacks:

1) Long set-up time caused by the need to attach the markers to each subject in the scene. That might also constrain the number of subjects.
2) Limiting the motion. This issue prevents mocap to be used to analyze the motion of the top athletes when they are producing their best performances.

To overcome this problem our research group has developed a technology, called Video Mocap (VMocap) [12], which is capable to reconstruct the motion from video. The method combines well-known triangulation techniques with recent development in the area of human pose detection [13], [14]. This technology enables us to record and reconstruct human motion data recorded in the human's natural setting without any interference.

In this paper, we describe how we successfully recorded a competitive team sports match in full and obtained data that was later processed offline using VMocap. This setup created several strict constraints:

1) No intervention on the subjects, so they can give their full performance in a competitive setting.
2) No interference with supporting operations, like TV

*This research was supported by JSPS KAKENHI 17H00766 (PI: Y. Nakamura) and NTT DOCOMO, INC..

[1] Y. Ikegami, M. Nikolić, A. Yamada, L. Zhang, N. Ooke and Y. Nakamura are with Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-Ku, Tokyo 113-8656, Japan {surname}@ynl.t.u-tokyo.ac.jp

[2] Milutin Nikolić is with the Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia milutinn@uns.ac.rs

[3] Lei Zhang is with School of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, 1 Zhanlanguan Road, Xicheng, 100044 Beijing, China leizhang@bucea.edu.cn

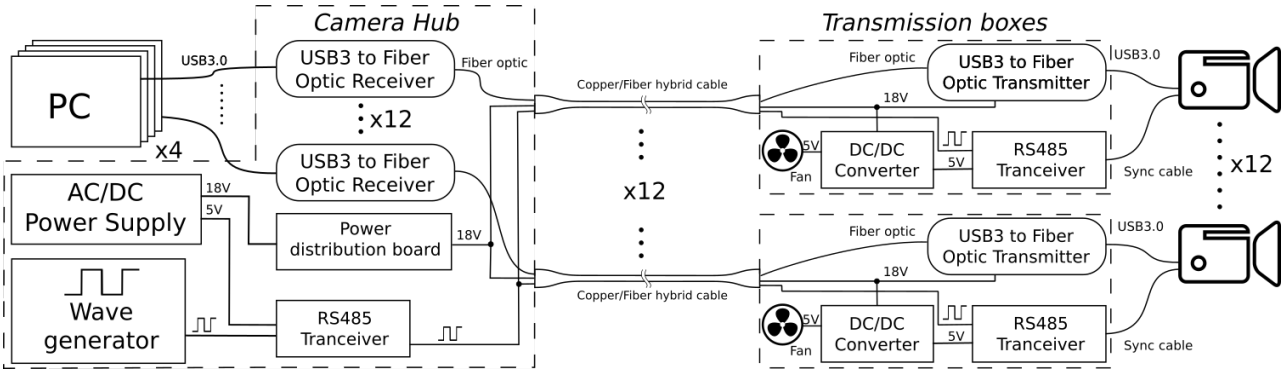


Fig. 1. Schematic of the camera network. Receiving side is depicted on the left, while the transmission side is on the right of the figure.

coverage, team staff, and audience.
3) Need for large recording volume, due to the large size of the court.
4) Need for synchronized high-speed cameras with large distances between them.
5) Required ability to store a large amount of incoming data in real-time.
6) A stable system with the ability to run for a prolonged duration is required.
7) Calibration procedure required for triangulation.

If any of the aforementioned requirements was not met, we either won't be allowed to record the data or the obtained data would be of no value of us. Addressing these points is a challenging task and in subsequent sections, we describe in detail how we addressed the enumerated issues.

II. CAMERA NETWORK ARCHITECTURE

VMocap uses RGB cameras as the only sensor, thus the performance of the camera network greatly influences the performance of the VMocap and the quality of the reconstructed motion. RGB cameras need to have:

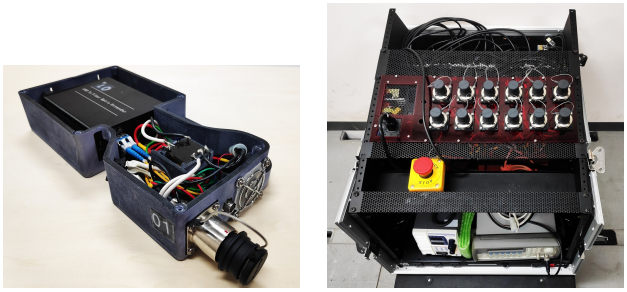
1) high resolution for high reconstructed motion accuracy,
2) high framerate, especially when dealing with fast motions, for smooth reconstructed motion
3) synchronized image capture for proper triangulation.

Modern USB3.0 cameras fulfill this requirement set, are relatively inexpensive and widely available.

The previous work with VMocap [12] was shown to be effective for a smaller area ($10m \times 10m$). Scaling that up to the size of the team sports playfield poses significant challenges:

1) RAW 2K images at 120Hz have to be transmitted to distances up to 200m.
2) Synchronization signal has to come from a single source and has to be transmitted over a long distance.
3) The camera may be far away from any power source.

These requirements led to a design of custom electronics, both on transmission (camera) side and reception (operation desk) side. The functional diagram of the implemented camera network is shown in Fig. 1.



(a) Transmission box (b) Camera hub

Fig. 2. Custom-made electronic hardware.

A. Data transmission

Transmitting RAW 2K images at 120Hz over long distances using multiple USB repeaters isn't a viable option. A large number of interconnections would compromise the robustness. Using Cat6 twisted-pair cables can provide required bandwidth, but are applicable only up to 100m [15]. For these reasons, we opted to use USB3.0 over fiber optic extender capable of transmitting the data up to 250m. It consists of a transmitter on the camera side, placed inside the transmission box (Fig. 2(a)), and USB3.0 dongle on the PC side, placed in camera hub as shown in Fig. 2(b).

B. Camera synchronization

To synchronize the cameras, image acquisition has to be triggered by an external signal. The 5V square signal used for triggering is generated by a wave generator at the operation desk. It is amplified by RS485 transceivers on both transmission and reception side (Fig. 1). Using this solution, the maximum distance of 1200m can be achieved at 100KHz, which is well above system requirements at 200m with 120Hz. On the transmission side, the output channel of RS485 transceiver is connected to the camera via synchronizing cable.

C. Power distribution

On the camera side, USB3.0 to fiber optic transmission box requires at least 12V for operation. Also, RS485 transceivers require 5V. Hence, the power at 18V is generated by a DC power supply at the operation desk and distributed

to cameras. Higher voltage was used to compensate for the voltage drop caused by long distribution lines. To obtain a stable power of 5V, a DC/DC converter was used on the camera side.

D. Implementation details

The camera hub was implemented to be able to connect 12 cameras, although in the experiments only 10 cameras were used. Instead of using 3 different cables (fiber optic for data, copper wires for RS485 and power) a single Copper/Fiber Hybrid Cable with two optical and 5 copper lines was used. That made system much cleaner and easier to setup. One spare copper line was used for common ground. Longest cables used were 150m long, and the voltage drop was around 3V. All camera side electronics was packed into a compact, self-sufficient transmission box shown in Fig. 2(a). The box includes a power switch, hybrid and synchronizing cable jack, single USB3.0 port, and a cooling fan. That box provided all the necessary ports and connections for the proper functioning of a single RBG camera with USB3.0 interface. When all 10 cameras were connected and running the system consumed around 130W.

III. COMPUTER ARCHITECTURE

To be able to acquire, process and save image data computer architecture has to be considered carefully. State of the art CPU AMD Ryzen Threadripper 2990WX with 32 cores, coupled with 128GB DDR4 RAM, ROG Zenith Extreme Alpha chipset and Windows 10 64-bit operating system were chosen.

The cameras are capturing RAW images at resolution 1920×1200 at 120Hz. Hence, each camera requires that 276MBps of raw data has to be received. Although USB3.0 specification states that the nominal signaling data rate of the physical layer is 625MBps, because of transmission overhead single USB controller can't receive data from two cameras at 120Hz. For the PC to accommodate four cameras, two StarTech PEXUSB314AV PCIe USB3.1 extension cards with two USB3.1 controllers each are added.

To accommodate the persistent storage for the high data throughput during a full match, the computers were equipped with four Intel SSD 660P 1.0TB M2 hard drives. The drives were further organized into a RAID0 array to boost the writing performance from 1.6GBps to 5.5GBps for sequential single thread and from 75MBps to 150MBps for random access single thread writing.

A. Data compression

Each incoming RAW image was 2.3MB, and saving them directly wasn't practical for recording long sports match because it required 16.5GB per camera per minute, thus some type of data compression had to be used. Sizes of images compressed to lossless PNG or lossy JPG format were around 2MB or 0.5MB respectively, which wasn't sufficient compression rate. Hence, the incoming images had to be compressed as a video, which created another set of constraints.

Because of the long recording duration, to avoid buffer overflow, each frame has to be compressed within the sampling period. To parallelize the process, thereby utilizing the high number of CPU cores, recording was split into 15sec video segments. This way multiple video segments can be compressed in parallel.

FFmpeg [16] was used to compress images to MPEG4 video at the bitrate of 65740Kbps (8217.5KBps), thus achieving a compression ratio of approximately 34:1. When a single camera is connected, the acquisition buffer stabilizes at around 1800 RAW images while creating two videos in parallel. In the case when all four cameras are connected, acquisition buffer stabilizes at 3400 RAW images per camera, while four videos in parallel per camera are created.

With the given setup, the hardware exceeded the required 32.8MBps writing speed, while CPU utilization was at around 60%. The acquisition system was able stably to run for more than an hour, without skipping a single frame from any camera.

IV. INTEGRATED CAMERA CALIBRATION

We make camera calibration for the whole game by integrating (a) intrinsic camera parameters, (b) measurements of feature points by surveying instruments and (c) bundle adjustments using markers. Four intrinsic parameters and five distortion parameters of each camera are obtained by OpenCV using a checkerboard after fixing its focal length. The process of calibration of extrinsic parameters, such as the position and orientation of each camera in the arena coordinates, is described in this section.

The relationship between the 3D positions in the arena space and the 2D position in the camera images is represented by

$$ {}^{i}\boldsymbol{Y}\,{}^{i}\boldsymbol{S} = {}^{i}\boldsymbol{A}\,{}^{i}\boldsymbol{B}\boldsymbol{X} \tag{1} $$

where $\boldsymbol{X} = (\boldsymbol{x}_1 \ldots \boldsymbol{x}_n)$, $\boldsymbol{x}_j = (x_j\ y_j\ z_j\ 1)^T$ is the j-th point in the arena space. ${}^{i}\boldsymbol{Y} = \left({}^{i}\boldsymbol{y}_1 \ldots {}^{i}\boldsymbol{y}_n\right)$, ${}^{i}\boldsymbol{y}_j = \left({}^{i}u_j\ {}^{i}v_j\ 1\right)^T$ is the 2D image position in camera i of the j-th point in the arena space. ${}^{i}\boldsymbol{S} = \mathrm{diag}\left\{{}^{\mathrm{i}}\mathrm{s}_1, \cdots, {}^{\mathrm{i}}\mathrm{s}_\mathrm{n}\right\}$ and ${}^{i}s_j$ is given by ${}^{i}s_j = (0\ 0\ 1)\ {}^{i}\boldsymbol{A}\,{}^{i}\boldsymbol{B}\,\boldsymbol{x}_j$. The two matrices of ${}^{i}\boldsymbol{A}$ and ${}^{i}\boldsymbol{B}$ are as follows:

$$ {}^{i}\boldsymbol{A} = \begin{pmatrix} {}^{i}f_x & 0 & {}^{i}c_x \\ 0 & {}^{i}f_y & {}^{i}c_y \\ 0 & 0 & 1 \end{pmatrix},\ {}^{i}\boldsymbol{B} = \left({}^{i}\boldsymbol{R}\ {}^{i}\boldsymbol{t}\right) \tag{2} $$

where ${}^{i}\boldsymbol{A}$ is the intrinsic parameters of camera i such as the focal lengths and the optical center in the image plane, ${}^{i}\boldsymbol{R}$ and ${}^{i}\boldsymbol{t}$ are the extrinsic parameters such as the orientation and position of camera i. Note that the lens distortion is also compensated using the intrinsic parameters, which is not explicitly shown in the above equation. ${}^{i}\boldsymbol{A}$ and the distortion parameters are determined for each camera. The rest is to identify the extrinsic parameters, namely ${}^{i}\boldsymbol{B}$.

Since the focus is fixed at a few tens of meters, it is ideal to use a large checkerboard, but not practical. One may suggest

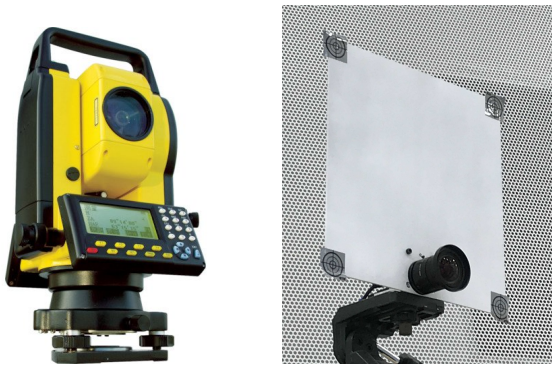


Fig. 3. Total station with the frilled camera

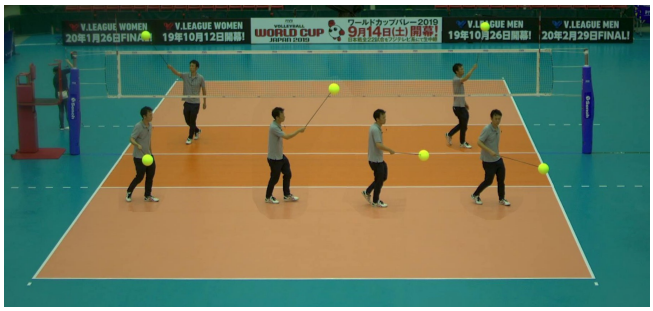


Fig. 4. Bundle adjustment data acquisition: Target ball sweeping the space

bundle adjustments, where one or a few markers like on a T-shaped wand are used. It requires to sweep almost the whole area by the markers. The arena space is difficult to sweep out, since the view angle of the camera that fits the width of the arena, results in very high recording volume. This results in limited sweep and inaccuracy of calibration.

To solve the problem we developed a method to combine direct measurements and bundle adjustment. For direct measurements, we use the total station, a geodetic surveying equipment that combines a theodolite and a laser range finder. We used STS-200s of Survey Techno-Science Inc. (Fig. 3), which has the accuracy within 20 seconds in both pan and tilt angles of the theodolite and within $\pm$1mm at 28.8m of the laser range finder. First, it is used to measure the 3D positions of the feature points such as the centers and corners of the courts. Second, the position and orientation of a camera are calculated by measuring the four corner markers of a square plate (40cm$\times$40cm) that is fixed to the camera at a set position with the optical axis perpendicular to the plate (Frilled camera, see 3.)

The extrinsic parameters are obtained by the following steps:

Step 1. Compute the position and orientation of the cameras from the measurements of the four corners of the plate.

Step 2. Set the 3D positions of the feature points as $\boldsymbol{X}_f \in R^{4\times n_f}$ (See Fig. 5). Set the unknown 3D positions of the target ball (Fig. 4) for the bundle adjustment as $\boldsymbol{X}_b \in R^{4\times n_b}$. Solve the intrinsic parameters, the extrinsic parameters and the unknown 3D positions of the target ball simultaneously by minimizing the following equation, where the initial

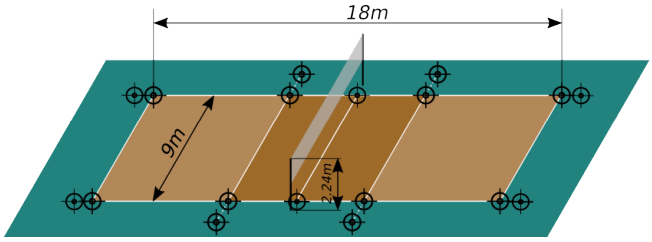


Fig. 5. Sketch of the court with all the feature points

values and the upper and lower bounds of the intrinsic and extrinsic parameters are set using the results of intrinsic parameter calibration and Step 1.

$$\min_{{}^i\boldsymbol{A},\,{}^i\boldsymbol{B},\,\boldsymbol{X}_b} \{\frac{k_f}{n_f}\sum_i \|\,{}^i\boldsymbol{Y}_f\,{}^i\boldsymbol{S}_f - {}^i\boldsymbol{A}\,{}^i\boldsymbol{B}\,\boldsymbol{X}_f\|^2 + \frac{k_b}{n_b}\sum_i \|\,{}^i\boldsymbol{Y}_b\,{}^i\boldsymbol{S}_b - {}^i\boldsymbol{A}\,{}^i\boldsymbol{B}\,\boldsymbol{X}_b\|^2\} \quad (3)$$

where the Frobenius norm is used, $k_f \geq 0$, and $k_b \geq 0$. We use a nonlinear optimization solver "lsqnonlin" with trust-region reflective method of MATLAB Optimization Toolbox for minimization. Note that the computation is longer when $k_b \neq 0$ since the second term involves a lot more variables to solve. One can first compute a solution by setting $k_b = 0$ and then search for another by setting $k_b > 0$.

V. EXPERIMENT

The recording took place during the international friendly volleyball game between women's national teams of Japan and Chinese Taipei. Two games were scheduled during two consecutive days. The event took place at Fukaya Big Turtle arena on August 10 and 11, 2019. The event had a large number of spectators and was covered live by a TV station.

A. Arena setup

For the recording, ten Basler acA1920-155uc cameras with USB3.0 interface were used. The cameras had to cover of 18m$\times$9m playing area with an additional 3m free zone around it, to a height of 3m. Four cameras were placed court-side on the first level (cameras no. 5,6,9 and 10), two cameras are placed in spectators stands on the second floor (cameras 7 and 8) and four cameras on 3rd-floor galley (cameras 1,2,3 and 4) were recording the full court. The furthest distance between two cameras (cameras 1 and 6) was around 60 meters. Cameras close to the playfield were equipped with wide-angle 8mm lenses (cameras 5 to 10), while cameras 1-4 were equipped with 21mm lenses.

Operations desk was on the court level behind the service lines. It consisted of four PCs connected to camera hub. Operation desk was in control of data recording software, camera trigger generator and system monitoring. The cameras were connected to the operation desks by 50m (cameras 3, 5, 9 and 10), 100m (cameras 4, 7 and 8) and 150m (cameras 1, 2 and 6) Copper/Fibre hybrid cables. The sketch

Fig. 6. Images from all 10 cameras. Row 1 contains images from cameras 1 to 5, row 2 contains images from cameras 6 to 10

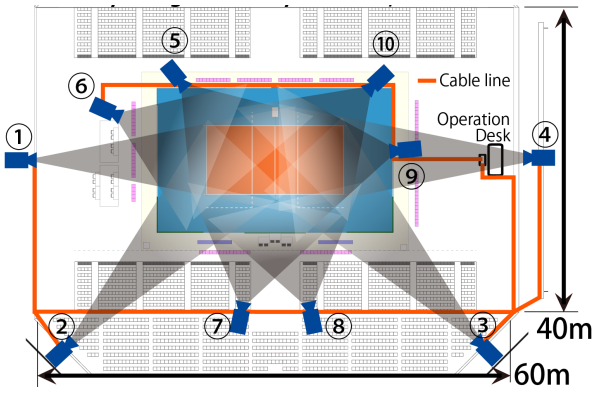


Fig. 7. Arena: Positions and FOVs of cameras and cable lines

of the arena with camera positions, operations desk and camera lines is shown in Fig. 7.

Avoiding camera movement was of paramount importance, so where it was available the cameras were clamped to the handrail. Courtside cameras were fixed on tripods. On the game day, each camera had to be guarded by a single person, ensuring that spectators don't move any camera by accident.

B. Recorded data

During the two-day event, not only the games were recorded. The teams training sessions and pre-game warm-up was recorded as well. The duration of recorded videos is summarized in table I. It can be seen that across two days, we have successfully recorded almost 6 hours of video at 120Hz totaling at 1.6TB of compressed video for all ten cameras. That amount of compressed data came as a result of processing 54TB of RAW image data.

Images obtained from all ten cameras at the same moment can be seen in Fig. 6. The big crowd and non-laboratory setting of the experiment can be noticed straight away. The cameras are primarily focused on the playing field. The experiment doesn't disturb any aspect of the game and the athletes were recorded while giving their best performances.

C. Reconstucted data

Using VMocap technology in conjunction with calibration data, the trajectories of the player's body-parts in the 3D space were reconstructed from recorded videos [12]. The

TABLE I

RECORDING DURATION AND DATA SIZE PER CAMERA

	Day One	Day Two
Training session I	97min	-
Training session II	56min	-
Warm up session	18.5min	26min
Set 1	22.5min	21min
Set 2	24min	21min
Set 3	29min	27.5min
Total min	247min	95.5min
Total GB	112 GB	47 GB

reconstructed trajectories was then used for two different purposes:

1) Single player spike analysis: In this case the focus was on bio-mechanical analysis of player's motion. Starting from the reconstructed motion of body parts, the skeleton model of a single player was reconstructed and muscle tensions were calculated [2], [19]. The musculosketetal model is generic and may have small errors due to individuality. The inertial parameters of each link were estimated from a statistical database and scaled uniformly to the weight and height of the athlete. The estimated muscle tensions are calculated so they reproduce recorded motion. The reconstructed musculoskeletal model was shown in Fig. 8

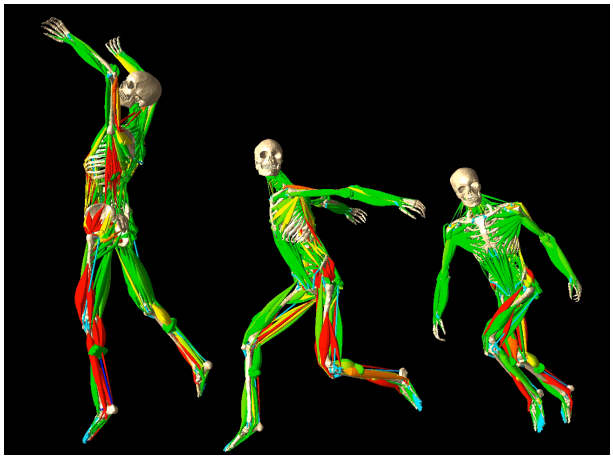
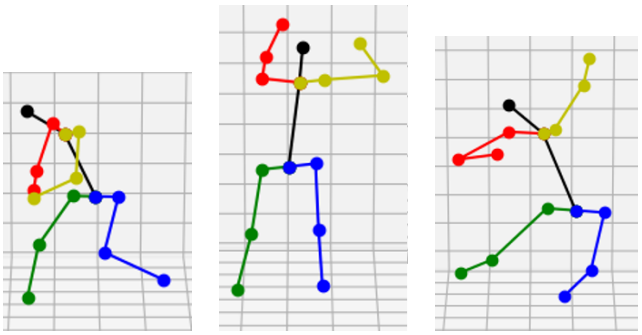


Fig. 8. Reconstructed muscle activation during spiking



(a) bump (b) set (c) spike

Fig. 9. Examples of extracted behaviors.

2) Behavioral analysis of the team: This time the focus was on study of players behavior and understanding of player's interaction. Hence, the whole team had to be reconstructed. For purpose of understanding motion only the skeletal model was sufficient [20]. Examples of extracted behaviors are shown in Fig. 9. The interaction of the players was investigated by correlating the time series of extracted behaviors.

The biomechanical and behavioral analyses of reconstructed motion are out of scope of this paper.

VI. CONCLUSION

In this paper, we have described how we have approached and successfully solved the problem of capturing the motion for long periods using only RGB cameras with high framerate. To do so we had to:

1) Solve the problem of synchronized long-distance high-bandwidth data transmissions by designing custom electronics and using USB3.0 to Fibre Optic extender.
2) Select appropriate computer architecture and write appropriate software capable of processing and storing incoming high-bandwidth image data.
3) Devise and employ calibration procedures that can be used to calculate camera positions with high accuracy. For that purpose, we have used geodetic surveying tools to measure initial camera poses. Those poses are later refined using bundle adjustment.
4) We used the developed system for video motion capturing of the international friendly match of the women's volleyball between the national teams of Japan and Chinese Taipei. The total RAW data of 54TB was provided from the ten cameras at 120FPS and compressed in real time and stored as 1.6TB video data.
5) Using previously developed VMocap technology, the full-body motion and/or muscle activations was reconstructed. The method is applicable to both single or multiperson reconstruction that can be used as a starting point for behavioral or biomechanical research.

ACKNOWLEDGMENT

We would like to thank the Japan Volleyball Association for providing us an opportunity to capture the data at the international friendly match of the senor women's national team of Japan. The capturing at the match would not be possible without the participation as capturing crew of Yan Huang, Taiki Ishigaki, Takahiro Nakanishi, Yuichi Sakemi, Akihiro Sakurai, Yoshihisa Shibata, Ko Yamamoto, Ryo Yanase, and Tianwei Zhang.

REFERENCES

[1] K. Yamane. Estimation of physically and physiologically valid somatosensory information. Proceedings of IEEE International Conference on Robotics and Automation, Barcelona, Spain, April 2005.
[2] A. Murai, K. Kurosaki, K. Yamane, and Y. Nakamura. Musculoskeletal-see-through mirror: Computational modeling and algorithm for whole-body muscle activity visualization in real time. Progress in Biophysics and Molecular Biology, 103(2):310317, 2010. Special Issue on Biomechanical Modelling of Soft Tissue Motion.
[3] W. Takano and Y. Nakamura. Synthesis of whole body motion with pose-constraints from stochastic model. IEEE International Conference on Robotics and Automation pages 18711876,2014
[4] N. F. Duarte, M. Raković J. Tasevski, M. I. Coco, A. Billard and J. Santos-Victor, Action Anticipation: Reading the Intentions of Humans and Robots, in IEEE Robotics and Automation Letters, vol. 3, no. 4, pp. 4132-4139, Oct. 2018.
[5] P. Schydlo, M. Raković, L. Jamone and J. Santos-Victor, Anticipation in Human-Robot Cooperation: A Recurrent Neural Network Approach for Multiple Action Sequences Prediction 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, QLD, 2018, pp. 1-6.
[6] M. Hehn and R. DAndrea, Real-Time Trajectory Generation for Quadrocopters, in IEEE Transactions on Robotics, vol. 31, no. 4, pp. 877-892, Aug. 2015.
[7] I. Kovačić, D. Radomirović, M. Zuković, B. Pavel, M. Nikolić, Characterisation of tree vibrations based on the model of orthogonal oscillations, vol. 8, no. 1, pp 8558-8568, June 2018.
[8] Motion Analysis coroporation, https://www.motionanalysis.com/
[9] Optitrack corporation, https://optitrack.com/
[10] Qualysis corporation, https://www.qualisys.com/
[11] XSens corporation, https://www.xsens.com/
[12] T. Ohashi, Y. Ikegami, K. Yamamoto, W. Takano and Y. Nakamura, Video Motion Capture from the Part Confidence Maps of Multi-Camera Images by Spatiotemporal Filtering Using the Human Skeletal Model, 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, 2018, pp. 4226-4231.
[13] Z. Cao, T. Simon, S.E. Wei and Y. Sheikh, Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, CVPR, 2017, Honolulu, USA
[14] K. Sun, B. Xiao, D. Liu and J. Wang, Deep High-Resolution Representation Learning for Human Pose Estimation, CVPR, 2019, Long Beach, USA
[15] Telecommunications industry association. Balanced twisted-pair telecommunications cabling and components standard (TIA-568.2-D) Retrieved from https://global.ihs.com/doc_detail.cfm?&csf=TIA&item_s_key=00339843
[16] FFMpeg multimedia framework. https://ffmpeg.org/
[17] USB 3.0 Promoter Group, Universal Serial Bus 3.2 Specification, USB Implementers Forum 2017, [Online]. Available: http://www.usb.org. [Accessed: Sep. 6, 2019].
[18] Z. Zhang, A Flexible New Technique for Camera Calibration, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 11, pp 1330-1334, Nov. 2000
[19] Y. Nakamura, K. Yamane, Y. Fujita and I. Suzuki, Somatosensory computation for man-machine interface from motion-capture data and musculoskeletal human model, in IEEE Transactions on Robotics, vol. 21, no. 1, pp. 58-66, Feb. 2005.
[20] K. Tsuzuki, W. Takano, and Y. Nakamura. Linguistic interpretation of human behavior by using motion symbol and corpus. In The Proc. of JSME annual Confer ence on Robotics and Mechatronics (Robomec) 2018, pp. 2P2-B18. The Japan Society of Mechanical Engineers, 2018.