High-resolution Performance Capture by Zoom-in Pan-tilt Cameras

Norimichi Ukita   Shigenobu Fujine   Norihiro Hagita
Nara Institute of Science and Technology
Japan
ukita@ieee.org

Abstract

We have developed a system with multiple pan-tilt cameras for capturing high-resolution videos of a moving person. This system controls the cameras so that each camera captures the best view of the person (i.e. one of body parts such as the head, torso, and limbs) based on criteria for camera-work optimization. For achieving this optimization in real time, time-consuming pre-processes, which give useful clues for the optimization, are performed in a training stage. Specifically, a target performance (e.g. a dance) is captured to acquire the configuration of the body parts at each frame. In a real capture stage, the system compares an online-reconstructed shape with those in the training data for fast retrieval of the configuration of the body parts. The retrieved configuration is used by an efficient scheme for optimizing a camera work. Experimental results show the camera work optimized in accordance with given criteria. A high-resolution 3D videos produced by the proposed system are also shown as a typical use of high-resolution videos.

1. Introduction

Performance capture[1, 2, 3, 4] is an important technique for 3D digital archiving and movie production. While the previous work on performance capture achieves fine 3D reconstruction and texture mapping, its accuracy has limitations if a target person is captured in low resolution, as shown in the left example in Fig. 1. Not only for performance capture but also for improving a variety of image understanding algorithms, high-resolution images and videos of the target person are useful. For example, detailed facial parts can be obtained for emotion understanding. Precise skills (e.g. theatrical forms, sports skills, and crafts) can be analyzed from the 3D motion of the person. This paper proposes a system for capturing high-resolution videos of a moving person by employing zoom-in pan-tilt cameras.

While a zoom-in camera can capture the high-resolution images of a part of a person, its field of view is limited and cannot contain the whole body of a person. This limitation
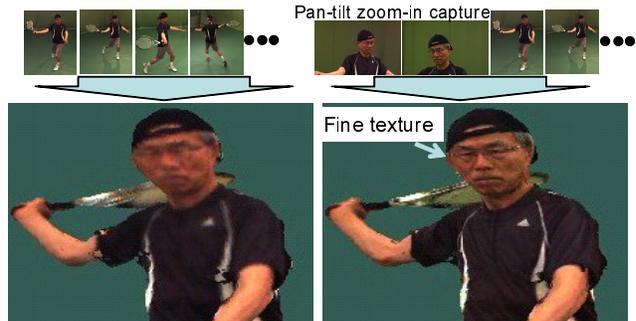

Figure 1. Comparison between conventional and proposed approaches. (Left) Conventional approach: only zoom-out images are captured by the conventional approaches. (Right) proposed approach: in addition to the zoom-out images, zoom-in images are captured continuously by pan-tilt cameras. (Bottom) 3D textured surfaces produced by the captured images demonstrate the effectiveness of the zoom-in images.

causes need for dynamic camera working (i.e. pan-tilt control) of multiple zoom-in cameras in order to continuously capture the whole body of the moving person. In an example shown in the right of Fig. 1, the head of a performer was captured by zoom-in pan-tilt cameras. These cameras enable high-resolution performance capture by employing existing methods[1, 2, 3, 4].

Technical issues in developing the system with zoom-in pan-tilt cameras are (1) criteria for a camera work taking into account the dynamic configuration of moving body parts and (2) real-time optimization of the camera work.

The proposed criteria are designed based on the priority of body regions and the quality of 3D reconstruction. The real-time optimization is achieved by reducing possible camera-works, offline training of possible human body configuration, and a simple optimization scheme for camera working.

2. Related Work

While a variety of camera parameters can be controlled in an active vision system as mentioned in [5], our work
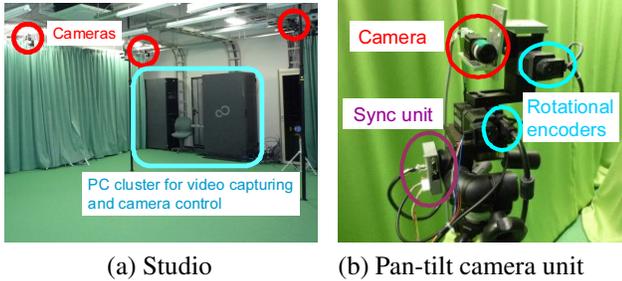
(a) Studio   (b) Pan-tilt camera unit

Figure 2. Experimental environment and system. (a) Multiple synchronized cameras and computers located in a chroma-key studio. (b) A pan-tilt camera unit consists of a camera, a computer-controllable pan-tilt unit, pan-tilt encoders for inquiring accurate pan-tilt angles, and a synchronization unit that gets the pan-tilt angles in synchronization with image capturing timing.

focuses on the pan-tilt angles of fixed cameras. In this work, a general control issue[6] is not discussed, but we consider how to determine the pan-tilt angles of all cameras so that they cooperatively capture a target. The pan-tilt angles should be determined to capture important components of the target depending on the task. In general tasks, main components captured by the cameras are the 3D shape and textures of the target. For example, human behavior[7, 8] and emotion[9] can be represented by the motions of the body shape and the facial textures, respectively.

Previously, for accurate shape reconstruction, camera viewpoints are optimized for a stationary object[10]. As well as a 3D shape, a texture is targeted for object recognition with viewpoint optimization[11]. In terms of moving objects, people[12] and their faces[13] are captured in surveillance scenarios. The limitation of these works is that their targets are only simple objects (e.g. a static object, the whole body, and only a part of the body). While multiple parts of the whole body are targeted by multiple pan-tilt cameras in [14], its computational cost in optimizing the camera work is huge (i.e. not in real time) and it has been tested only in a simulation environment. For real performance capture, we propose real-time camera-work optimization in this paper.

3. System Architecture

The system consists of synchronized cameras and computers, which capture and analyze images online and control the cameras taking into account the analysis results. The cameras are positioned so that they surround a subject located around the center of a studio, as shown in Fig. 2 (a).

In our system, the focal length of every camera is fixed. Several cameras are able to control their pan-tilt angles and are possessed of a zoom-in lens for high-resolution capture, while other cameras are fixed and zoomed out for stably observing the whole body of a moving subject.

As shown in Fig. 2 (b), each pan-tilt camera allows us to obtain the pan-tilt angle synchronized with every image capturing timing. This function is crucial for reconstructing a 3D human body while the pan-tilt angle is controlled dynamically. This is because external camera parameters, which are required for 3D reconstruction, can be computed from those in the canonical camera angle and the correct pan-tilt angle at each moment[15].

All computers with the cameras send captured images to a server computer that reconstructs a 3D shape and determine a camera work at each time.

4. Body Parts Retrieval

In the previous camera-work optimization system[14], all continuous pan-tilt angles (e.g. 0.1 degrees, 0.01 degrees) are evaluated. This requires huge computational cost due to combinatorial explosion. For reducing the computational cost, we evaluate only a limited number of pan-tilt angles for real-time optimization. The pan-tilt angles should be limited carefully in order to search for a nice camera work. Specifically, the pan-tilt angles are limited so that each camera gazes at one of body parts (e.g. head, torso, and limbs). This limitation reduces a search space while any body region can be possibly captured in high resolution.

4.1. Offline Training

The camera working for body parts requires segmenting a human body into the body parts. Real-time 3D body-part segmentation can be achieved with training data of real images[16] and simulation images[17]. In contrast to simulation images[17], real images[16] have the advantages in terms of accuracy and flexibility for complex shape and motion of a target. This method[16], however, needs specially-colored clothing, where each body part is colored differently, for training images. To alleviate this labor and cost, easier segmentation is proposed in this paper.

For training, the system captures a subject whose performance and clothing are similar to those in a real capture stage. Any color and textures of the clothing are acceptable. That is, the 3D shape of the subject must be similar between training and test data.

Figure 3 shows the abstract of the proposed body-parts labeling. The whole body is divided into five parts, namely the head, torso, legs, right arm, and left arm. Instead of the colored clothing, only colored gloves are employed. In each image, pixels corresponding to the hands and the face are labeled by color detection and face detection[18], respectively. A surface point cloud is reconstructed by shape-from-silhouettes[19].

The pixelwise labels of the hands and the face are projected onto the reconstructed 3D points and regarded as their initial seeds in the 3D body. The initial seeds of the
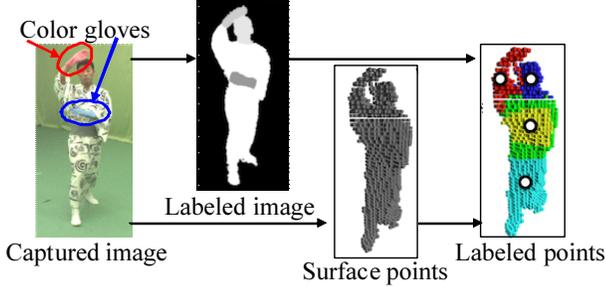
Figure 3. Labeling 3D body parts in reconstructed voxels. Each pixel in a labeled image has one of the following five labels: background, head, left hand, right hand, and other body regions. Surface points are reconstructed from multiview silhouettes. Each surface point is labeled by one of the following five parts, namely the head, torso, legs, right arm, and left arm. Small circles superimposed in the labeled points depict the centroids of the five parts.

torso and the legs are given in the median and 30cm-above-bottom of the reconstructed points, respectively. Then unlabeled points are labeled based on the nearest neighbor labeled points. Since the proposed system controls camera views towards the centroids of body parts in a real capture stage, the centroid of the labeled points is computed in each body part in a training stage. The centroids are recorded as training data with the surface points of the whole body.

The training data described above is obtained in all frames of a target human-motion sequence, which is captured in advance for training.

4.2. Online Retrieval

In a real capture stage, the surface point cloud, s, is reconstructed by shape-from-silhouettes. The centroids of all body parts in s are obtained for camera working. Note that, in the online capture, a target person wears no colored gloves. The centroids are obtained by retrieving the s's nearest neighbor from the training data.

For nearest neighbor retrieval of a human body shape, while dense pointwise matching[20, 21] is robust, its computational cost is unsuited for real-time online retrieval. For efficient and robust retrieval of the nearest neighbor, a high-dimensional point cloud is translated to its volume descriptor[22], where the point cloud is divided into several bins. In our experiments, 1) the shape of the bin model is a cylinder, whose vertical center axis passes through the median of all points, 2) the cylinder is divided by 35 height divisions, 12 azimuth divisions, and 3 radius divisions, and 3) the entity of each bin is the number of points. The dimension of a descriptor is 1260. Dissimilarity between two descriptors, f_1 and f_2, is expressed by $||f_1 - f_2||^2$.
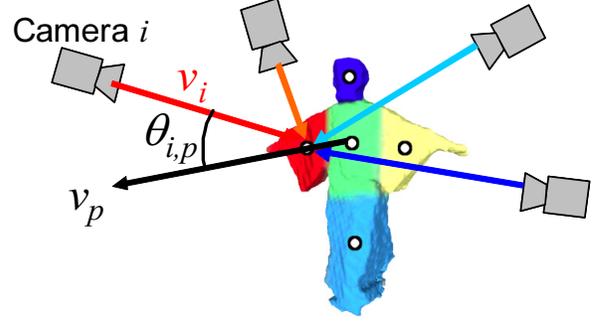

Figure 4. Utility function 1 for camera-work optimization: Occlusion check. Camera i might be able to capture the right arm because it is located between i and the body. In this example, vectors from i and the torso to the right arm are compared. Vectors with warmer colors indicate the view lines having higher scores.
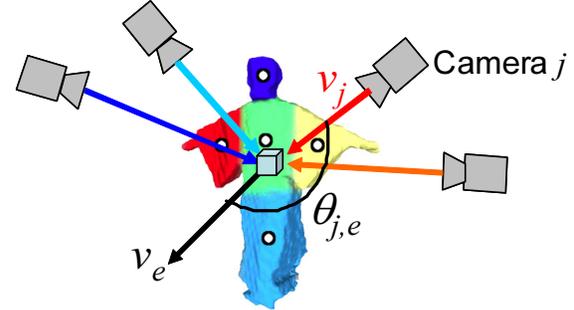

Figure 5. Utility function 2 for camera-work optimization: Reconstruction error reduction. Camera view lines to potential error regions (i.e. a cube in the figure) are evaluated for carving check. Vectors with warmer colors indicate the view lines having higher scores.

5. Real-time Camera-work Optimization

5.1. Utility Functions

A camera work is optimized so that the utilities of all cameras for capturing their target parts are maximized. The following utility functions are used for this optimization:

Occlusion avoidance (Fig. 4): A camera should observe its target part with no occlusion. For this occlusion avoidance, given a 3D vector from each camera i to the torso (denoted by v_i) and one more 3D vector from the torso to a potential target part p (denoted by v_p), angle $\theta_{i,p}$ between v_i and v_p is evaluated. The smaller $\theta_{i,p}$ is, the higher the utility of camera i becomes. This is because it seems very possible that part p is occluded by the torso if $\theta_{i,p}$ is larger (i.e. be close to π), in other words, if p is located beyond the torso from i's side. The utility function, $F_o(i,p)$, is expressed as follows:

$$F_o(i,p) \quad = \quad \cos \frac{\theta_{i,p}}{2} \qquad (1)$$

Reconstruction error reduction (Fig. 5): For improving a reconstructed shape, the pan-tilt cameras should be controlled for reducing reconstruction error. To this end, we focus on the property of shape-from-silhouettes, which carves false-positive error regions. Our strategy is that 1) whether or not each false-positive of a visual hull, which is obtained by shape-from-silhouettes, contained in training data can be carved is probed in a training stage and 2) then, in a real capture stage, the nearest neighbor of a reconstructed visual hull is searched for from the training data in order to retrieve false-positives that can be carved. While probing whether or not each false-positive can be carved needs huge computational cost, reconstructing a visual hull and its nearest neighbor search can be performed in real time.

Assume that the false-positives in a visual hull at each training frame are known by comparing the visual hull, h, and a more correct 3D shape, s, reconstructed by a slow but more accurate algorithm such as [23]; $(h - s)$ is regarded as the false-positives. Then view lines that can carve each false-positive with no occlusion are found. These view lines are found by checking whether or not each view line passing through a false-positive of interest intersects the surface of h[1]; if the view line intersects the surface, in other words, if the surface occludes the view line, this view line cannot carve the false-positive. The positions of the false-positives with such view lines in each training frame are recorded with their respective visual hull as training data.

In the online capture stage, body-shape retrieval gives us the positions of false-positives with view-lines for carving them. Let e and l_e denote a false-positive and a view-line that can carve it, respectively. The more a view-line from camera j to e (depicted by l_j in Fig. 5) is parallel to l_e, the more likely e can be carved by camera j. This property provides us the following utility function:

$$F_r(j, e) = \cos\theta_{j,e} \qquad (2)$$

Camera motion dynamics: Camera-work at time t should be determined so that it is close to camera-work at $t - 1$. This is because 1) fast camera motion causes motion blur in an image and 2) it is impossible to control the pan-tilt angle beyond the mechanical limitation on rotational velocity. In the proposed system, the upper limitation of a pan-tilt velocity is determined as follows. First of all, images are captured with various rotational velocities under actual camera parameters (e.g. shutter speed) and environment (e.g. lighting condition). Then the fastest velocity, v^{max}, which allows a camera to capture an unblurred image is selected[2] for thresholding in the following camera-work

---

[1]The surface was reconstructed by surface meshing[27].

[2]More practically, the acceleration of the pan-tilt angle should be controlled. In experiments shown in this paper, however, simple velocity thresholding was sufficient to capture unblurred images because the velocity was slow.
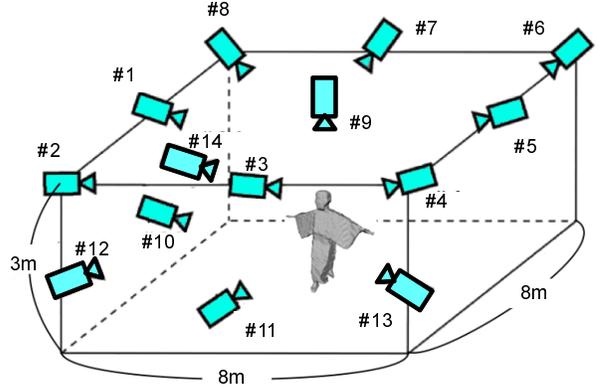

Figure 6. Camera setup in a studio. Camera number IDs correspond to those in Figure 8.

optimization.

5.2. Optimization Scheme

With utility functions (1) and (2), a camera work is optimized efficiently based on a greedy algorithm approach:

1. For all combinations of unoptimized cameras and parts, utility functions (1) and (2) are computed. Except that the combination of camera i and part p is neglected if the pan-tilt velocity from the current pan-tilt of i to the pan-tilt towards p is above v^{max}.

2. Given weight variables w_o and w_r, the following score is computed for all the combinations described above:

$$w_o F_o(i, p) + \frac{w_r}{|\boldsymbol{E}_{i,p}|} \sum_{e \in \boldsymbol{E}_{i,p}} F_r(i, e), \qquad (3)$$

where $\boldsymbol{E}_{i,p}$ denotes a set of false-positive voxels that are within the field of view of i-th camera when its view direction is towards p-th part. $w_o = w_v$ in our experiments.

3. In addition, the priority of each part, $w(p)$, is given to the score:

$$w_p \left(w_o F_o(i, p) + \frac{w_r}{|\boldsymbol{E}_{i,p}|} \sum_{e \in \boldsymbol{E}_{i,p}} F_r(i, e) \right), \qquad (4)$$

4. A combination with the highest score is selected. Then the camera work of the selected camera is determined. If any cameras remain to be unselected, go back to 1.

6. Experiments

All experiments were conducted in a studio shown in Figs. 2 (a) and 6. While the number of cameras used for
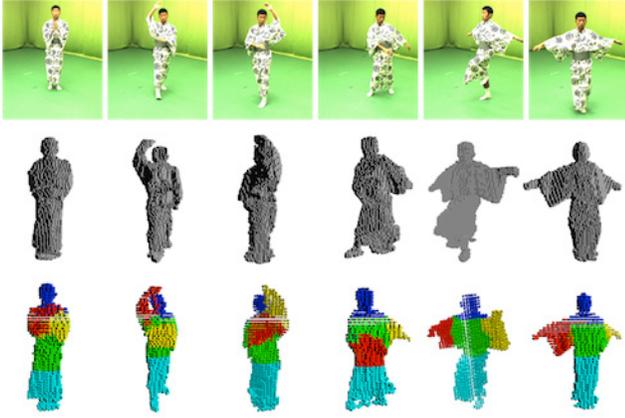
Figure 7. Top: image sequence obtained by a zoom-out camera. Middle: Online-reconstructed point cloud. Bottom: Retrieved body-parts labels.

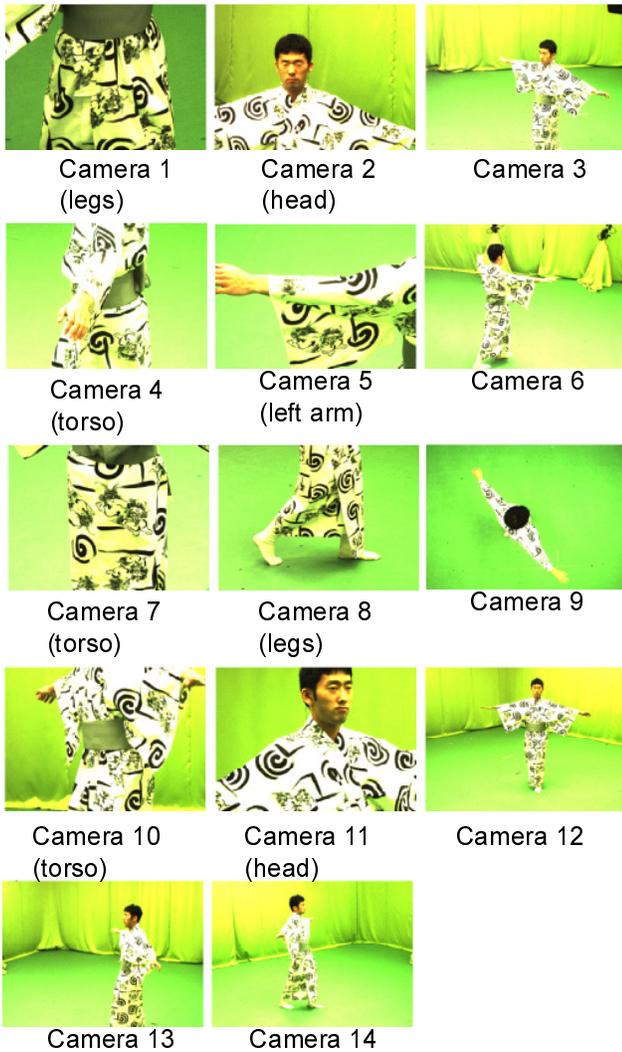| Camera 1 (legs) | Camera 2 (head) | Camera 3 |
Camera 4 (torso)	Camera 5 (left arm)	Camera 6
Camera 7 (torso)	Camera 8 (legs)	Camera 9
Camera 10 (torso)	Camera 11 (head)	Camera 12
Camera 13	Camera 14	

Figure 8. Six and eight images were captured by fixed zoom-out cameras (i.e. cameras 3, 7, 9, 12, 13, and 14) and zoom-in pan-tilt cameras (i.e. cameras 1, 2, 4, 5, 6, 8, 10, and 11).

performance capture was different among the experiments, a combination of zoom-in and zoom-out cameras was employed in all the experiments. This is because the zoom-out cameras were useful for continuously reconstructing the whole body that is required for body-parts retrieval. The zoom-in and zoom-out cameras, each of whose resolution was 1024×768 pixels, were possessed of lenses with 25mm and 6mm focal lengths, respectively.

In first experiments, body-parts retrieval was verified. Six zoom-out cameras were used for reconstructing the visual hull of a subject. 600 frames, which contained three-times repeated dance motion, were captured for training. Their volume descriptors were stored as a KD-tree for efficient search. The same subject was captured in a query sequence. The subject performed in the training and query sequences as similar as possible. Figure 7 shows the query image sequences captured by one of the cameras, the surface point cloud of the reconstructed visual hull, and the retrieved body-parts labels. It can be seen that the labels were obtained with no major mistakes. For applying the training data to anybody's dance similar to the training one, generality of the training data should be improved. The generality can be improved, for example, by probabilistic dimensionality reduction as achieved in [24, 25, 22].

Next, in addition to the six zoom-out cameras, the system used eight zoom-in pan-tilt cameras that were controlled by the proposed camera-work optimization using the retrieved body-part. The body-part retrieval was achieved with training data obtained in the first experiments described above. For validating the effects of the basic properties in the proposed optimization (i.e. computational cost and camera work based on the priorities of body parts), a stationary mannequin was captured. 14 images captured by the system are shown in Fig. 8. Images captured only by the zoom-out cameras were used for reconstructing a shape, which was employed for online body-parts retrieval.

The computational costs were 50ms in 3D reconstruction with $2cm^3$ voxels and less than 10ms for retrieval and camera-work optimization. The bottleneck in 3D reconstruction was image transmission by UDP and must be sped up for capturing a fast motion. On the other hand, the proposed retrieval and optimization could work in real time.

Since higher priorities were given to the head, torso, and arms, the zoom-in cameras captured the four parts evenly, as shown in Fig. 8. In this camera work, the theoretical spatial resolutions of the parts captured in high-resolution and the legs, which were captured only by the zoom-out cameras, were 0.7mm/pixel and 2.5mm/pixel, respectively.

For demonstrating an application of high-resolution videos, textured 3D mesh sequences were produced. Since accurate external camera parameters are required at each frame, the pan-tilt angle of each camera was controlled to either of several pre-calibrated pan-tilt angles, which was

Figure 9. Swinging arms: 12 captured images used for generating the righthand result in Fig. 10. 8 and 4 images were zoom-out and zoom-in images, respectively. In the zoom-out images, a human body region is cropped for conserving space.

Figure 10. Swinging arms sequence: results of mesh reconstruction and texture mapping.

Figure 11. Swinging arms sequence: results of mesh reconstruction and texture mapping obtained only from 14 zoom-out images.

Figure 12. 12 captured images used for generating the righthand result in Fig. 13.

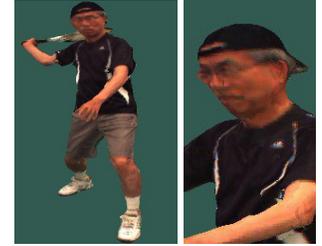Figure 13. Textured 3D mesh obtained from high-resolution images.

Figure 14. Textured 3D mesh obtained only from 12 zoom-out images.

Figure 15. 3D mesh obtained from high-resolution images.

Figure 16. 3D mesh obtained only from 12 zoom-out images.

the nearest neighbor of the pan-tilt angle optimized by the system, as proposed in cell-based tracking[26].

After capturing image sequences in a real capture stage, a point cloud was reconstructed by fusion of shape-from-silhouettes[19] and multiview stereo[23]. The point cloud was then used for surface meshing[27]. Texture mapping was implemented by simple OpenGL functions.

Figures 9 and 10 show captured synchronized images and textured mesh surfaces. In the experiments, eight zoom-out and four zoom-in cameras were used. The score function was designed to give high priority to the head for capturing high-resolution facial expression. For comparison, textured mesh surfaces produced only from zoom-out im-

ages are shown in Fig. 11. Superiority of the results acquired by the proposed system is apparent.

One more experiment was conducted for producing a textured mesh sequence. Images captured by the proposed system are shown in Fig. 12. High priority was given to a tennis racket as well as the head. The region of the racket was labeled by appearance-based object detection[18]. Textured mesh surfaces produced from all images captured by the proposed system and only from zoom-out images are shown in Figs. 13 and 14, respectively. It can be seen that the proposed system could get finer textures of the face.

For verifying the reconstructed shapes, non-textured mesh surfaces are shown in Fig. 15 and 16, which were reconstructed by employing zoom-in and only zoom-out cameras, respectively. It is also clear that a more smooth and accurate facial mesh was reconstructed by the zoom-in images; facial parts (i.e. nose and eyes) are seen in Fig. 15.

7. Concluding Remarks

This paper proposed the high-resolution video capture system with zoom-in pan-tilt cameras. As well as the zoom-in cameras, zoom-out cameras also compose the system for

continuously capturing the whole body of a target person. The shape of the whole body is employed for retrieving body parts, which are targeted by the zoom-in pan-tilt cameras. Several experiments demonstrated the elemental components of the proposed system (i.e. real-time body-parts retrieval and camera-work optimization) and applicability to producing high-quality 3D videos.

Future work includes real-time camera control, accurate camera parameter acquisition while dynamically controlling pan-tilt cameras[15], and developing a variety of applications of the proposed system.

The source codes of PMVS[23] and Poisson surface reconstruction[27] were given by Y. Furukawa and M. Kazhdan, respectively. The authors would like to thank all of them.

References

[1] D. Vlasic, I. Baran, W. Matusik, and J. Popovic, "Articulated Mesh Animation from Multi-view Silhouettes," *ACM Transactions on Graphics*, Vol.27, No.3, 2008. 1

[2] E. de Aguiar, C. Stoll, C. Theobalt, N. Ahmed, H.-P. Seidel, S. Thrun, "Performance capture from sparse multi-view video," *ACM SIGGRAPH*, 2008. 1

[3] J. Starck and A. Hilton, "Model-based human shape reconstruction from multiple views," *CVIU*, Vol.111, No.2, pp.179–194, 2008. 1

[4] A. Zaharescu, E. Boyer, and R. Horaud, "Topology-Adaptive Mesh Deformation for Surface Evolution, Morphing, and Multiview Reconstruction," *IEEE Trans. on PAMI*, Vol.33, No.4, pp.823–837, 2011. 1

[5] A. Ilie, Greg Welch, and M. Macenko, "A Stochastic Quality Metric for Optimal Control of Active Camera Network Configurations for 3D Computer Vision Tasks," *Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications*, 2008. 1

[6] P. Y. Oh and P. K. Allen, "Design of a partitioned visual feedback controller," *ICRA*, 1998. 2

[7] M. B. Holte, T. B. Moeslund, N. Nikolaidis, and I. Pitas, "3D Human Action Recognition for Multi-view Camera Systems," *3DIMPVT*, 2011. 2

[8] M. Grundmann, F. Meier, and I. A. Essa, "3D Shape Context and Distance Transform for action recognition," *ICPR*, 2008. 2

[9] B. Hernandez, G. Olague, R. I. Hammoud, L. Trujillo, and E. Romero, "Visual learning of texture descriptors for facial expression recognition in thermal imagery," *CVIU*, Vol.106, No.2–3, pp.258–269, 2007. 2

[10] W. Niew and M. Steinmetz, "Camera viewpoint control for the automatic reconstruction of 3D objects," *ICIP*, 1996. 2

[11] J. Denzler and C. M. Brown, "Information Theoretic Sensor Data Selection for Active Object Recognition and State Estimation," *IEEE Trans. on PAMI*, Vol.24, No.2, pp.145–157, 2002. 2

[12] F. Qureshi and D. Terzopoulos, "Surveillance in Virtual Reality: System Design and Multi-Camera Control," CVPR, 2007. 2

[13] T. Kato, Y. Mukaigawa, and T. Shakunaga, "Cooperative Distributed Tracking for Effective Face Registration," *MVA*, 2000. 2

[14] Y. Sofiane, N. Ukita, and M. Kidode, "An Assignment Scheme to Control Multiple Pan/Tilt Cameras for 3D Video," *Journal of Multimedia*, Vol.2, Issue 1, pp.10-19, 2007. 2

[15] T. Matsuyama and N. Ukita, "Real-Time Multi-Target Tracking by a Cooperative Distributed Vision System," *Proc. of the IEEE*, Vol.90, No. 7, pp.1136–1150, 2002. 2, 7

[16] N. Ukita, R. Tsuji, and M. Kidode, "Real-time Shape Analysis of a Human Body in Clothing with Time-series Part-labeled Volumes," *ECCV*, 2008. 2

[17] J. Shotton, A. Fitzgibbon, M. Cook, and A. Blake, "Real-time Human Pose Recognition in Parts from Single Depth Images," *CVPR*, 2011. 2

[18] P. A. Viola and M. J. Jones, "Robust Real-Time Face Detection," *IJCV*, Vol.57, No.2, pp.137–154, 2004. 2, 6

[19] G. Cheung, T. Kanade, J. Bouguet, and M. Holler, "A real time system for robust 3D voxel reconstruction of human motions," *CVPR*, 2000. 2, 6

[20] J. Starck and A. Hilton, "Correspondence labelling for wide-timeframe free-form surface matching," *ICCV*, 2007. 3

[21] A. Sharma, R. Horaud, J. Cech, and E. Boyer, "Topologically-robust 3D shape matching based on diffusion geometry and seed growing," *CVPR*, 2011. 3

[22] N. Ukita, M. Hirai, and M. Kidode, "Complex Volume and Pose Tracking with Probabilistic Dynamical Models and Visual Hull Constraints," *ICCV*, 2009. 3, 5

[23] Y. Furukawa and J. Ponce, "Accurate, Dense, and Robust MultiView Stereopsis," *IEEE Trans. on PAMI*, Vol.32, No.8, pp.1362–1376, 2010. 4, 6, 7

[24] N. D. Lawrence, "Probabilistic non-linear principal component analysis with Gaussian process latent variable models," *Journal of Machine Learning Research*, Vol.6, pp.1783–1816, 2005. 5

[25] J. M. Wang, D. J. Fleet, A. Hertzmann, "Gaussian Process Dynamical Models for Human Motion," *PAMI*, Vol.30, No.2, pp.283–298, 2008. 5

[26] T. Yamaguchi, S. Nobuhara, and T. Matsuyama, "Cell-based Object Tracking Method for 3D Shape Reconstruction Using Multi-viewpoint Active Cameras," *The 9th IEEE International Workshop on Visual Surveillance*, 2009. 6

[27] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson Surface Reconstruction," *SGP*, 2005. 4, 6, 7