

Digital Special Makeup by Onset Motion Capture

Jong-Sung Kim, Jae-Hean Kim

Visual Content Research Department, ETRI, South Korea

js.kim@etri.re.kr, gokjh@etri.re.kr

Abstract— In this paper, we present a new digital special makeup method to compose camera-captured human face or body images with CG ones for visual effects. Our method combines an active motion capture technique with an outside-in camera tracking technique to simultaneously track the 3D motion of both human and camera. Our method can be applied to digital special makeup in outdoors as well as indoors. The performance of our method has been demonstrated with various real face and body special makeup results.

Keywords— Special makeup, motion capture, camera tracking, visual effects

I. INTRODUCTION

Digital special makeup is a technique to compose camera-captured human face or body images with computer-generated (CG) ones for visual effects in TV and film production. Many computer vision and graphics techniques such as motion tracking and photo-realistic rendering should be integrated to create realistic digital special makeup effects. The 3D face or body motion tracking is essential to animate CG face or body. In addition, the 3D camera motion should be also tracked with high accuracy when the face or body image is captured by a moving camera. However, the simultaneous human and camera motion tracking as in digital special makeup has been studied not much and it still remains as a difficult problem in computer vision and graphics [1].

Motion capture records the human motion with specific sensors such as cameras or gyros and then maps the recorded motion data to CG models. Although motion capture has been introduced in computer animation [2], it is widely used for various applications in many different areas such as industrial training [3]. In addition, it can be used as a high-performance solution for 3D human face or body motion tracking [4, 5], which explains why we adopt a motion capture technology for digital special makeup. Although it is easy to solve the human motion tracking with motion capture, it is

not straight-forward to apply motion capture to the simultaneous human and camera tracking problem, particularly for digital special makeup in an outdoor environment.

In this paper, we present a new onset motion capture-based digital special makeup method to compose camera-captured human face or body images with CG ones for visual effects in TV and film production. Our method combines an active optical motion capture technology with an outside-in camera tracking technique to track the 3D motion of both human and camera onset with high accuracy, which enables us to apply our method to the digital special makeup of human face or body in outdoors as well as indoors contrary to our previous work on CG/real image composition [6].

The organization of this paper is as follows. Section 2 briefly describes our onset motion capture-based digital special makeup system. Section 3 explains the outside-in camera tracking technique in our method. Experimental results are given in Section 4, and our conclusions are drawn in Section 5.



Figure 1. Onset motion capture-based digital special makeup system

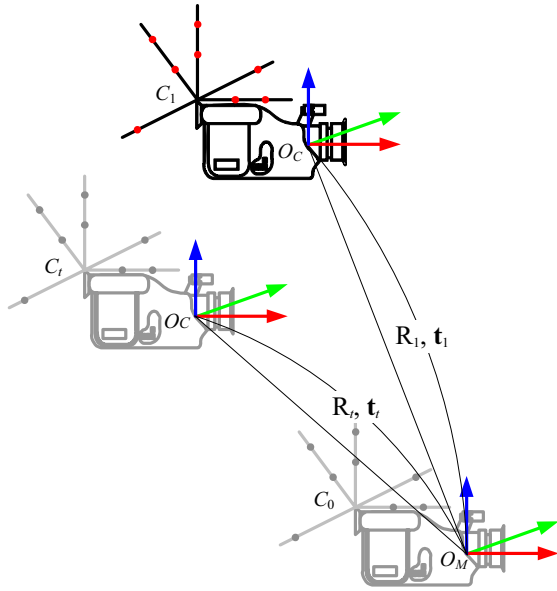


Figure 2. Relationship between the motion capture and the camera coordinate system for the camera with a camera target

II. DIGITAL SPECIAL MAKEUP

Our onset motion capture-based digital special makeup system (Fig. 1) is equipped with an active motion capture system and a HD video camera. As an active motion capture system, we employed Impulse system [7], which can perform motion capture in outdoors as well as indoors by using active LED markers and an advanced high-speed (up to 480 fps) imaging technology. The HD video camera records human face or body motion images of 1920×1080 pixels at 30 fps. A time-code reader is used to synchronize the HD video camera with the motion capture software in the host PC of Impulse system. For outside-in camera tracking, of which details will be explained in the following section, a camera target with eight active LED markers is attached to the HD video camera.

III. OUTSIDE-IN CAMERA TRACKING

In most camera tracking techniques, e.g., [6], the camera motion is observed with the camera itself (inside-out). However, in outside-in camera tracking techniques, the camera motion is observed with some external (outside-in) devices such as position sensors or other cameras. The main advantage of outside-in camera tracking is that it does not require to process camera-captured images, which makes it possible to track the camera motion

in real time or to perform camera tracking in a highly dynamic environment such as a motion capture studio.

Let us denote O_M and O_C the motion capture and the camera coordinate system, respectively, as in Fig. 2. At each time t , we obtain the 3D coordinates of N active LED markers on the human, $X_t = \{X_{t1}, \dots, X_{tN}\}$, and M ones on the camera target, $C_t = \{C_{t1}, \dots, C_{tM}\}$, in the x, y and z axes of the motion capture system. The 3D camera motion can be represented with a 3×3 rotation matrix R_t and a 3×1 translation vector t_t , which are the quantities that we should compute to compose the camera-captured image with CG one at each time.

Assume that we know $C_0 = \{C_{01}, \dots, C_{0M}\}$, the 3D coordinates of N active LED markers on the camera target when the 3D camera motion is null, i.e., $R_0 = I$ and $t_0 = 0$. Then, the 3D camera motion R_t and t_t of time t can be estimated from the relationship given by

$$C_{0m} = R_t C_{tm} + t_t. \quad (1)$$

The problem of estimating the transformation, including the rotation and the translation, between two coordinate systems is referred to as that of absolute orientation. Given C_0 and C_t , a closed-form solution of the rotation R_t be obtained by using unit quaternions [8] or orthonormal matrices [9]. The translation t_t is easy to determine once the rotation R_t is known. The disadvantage of this approach is that we should have both C_0 and C_t in advance. We can obtain C_t directly from the motion capture data of time t . However, we cannot obtain C_0 in the same way. We solve this problem as follows. At $t=1$, we estimate R_1 and t_1 by conducting the motion capture-based camera calibration

$$\hat{\Theta}, \hat{R}_1, \hat{t}_1 = \min \sum_{l=1}^L \|w_{1l} - h(W_{1l}, \Theta, R_1, t_1)\|^2, \quad (2)$$

where $W_1 = \{W_{01}, \dots, W_{0L}\}$ and $w_1 = \{w_{01}, \dots, w_{0L}\}$ denote the 3D and 2D image coordinates of active LED markers on our camera calibration wand (Fig. 1), respectively, Θ all of the internal camera parameters, including focal length, principal point, and lens distortion coefficients [10], and h the

projection function from the 3D space to the 2D image space [11]. Then, we compute C_0 with

$$\hat{C}_{0m} = \hat{R}_1 C_{1m} + \hat{t}_1. \quad (3)$$

As described above, we do not need to process camera-captured images in our outside-in camera tracking method. However, the camera-captured images can be used to improve the performance of our method. If the 2D coordinates of some of active LED markers on the human, of which the selection can be guided by a user, are obtainable at time t , the estimate of R_t and t_t can be refined by minimizing the following cost function

$$\hat{R}_t, \hat{t}_t = \min \left(\sum_{k=1}^K \|s_{tk} - h(S_{tk}, \Theta, R_t, t_t)\|^2 + \sum_{m=1}^M \|c_{0m} - h(C_{tm}, \Theta, R_t, t_t)\|^2 \right), \quad (4)$$

where $S_t = \{S_{t1}, \dots, S_{tK}\}$ and $s_t = \{s_{t1}, \dots, s_{tK}\}$ denote the 3D and 2D image coordinates of the user-selected, active LED markers on the human, respectively, and $c_0 = \{c_{01}, \dots, c_{0M}\}$ the 2D image coordinates of the active LED markers on the camera target. Note that $c_0 = c_1 = \dots = c_t$ due to the constraint in (1). The above step is called guided error minimization. The Gauss-Newton method [12] is adopted to minimize the cost function.

Figure 3 shows the dataflow of our digital special makeup method. The details of the 2D tracking and the image composition are described in [6].

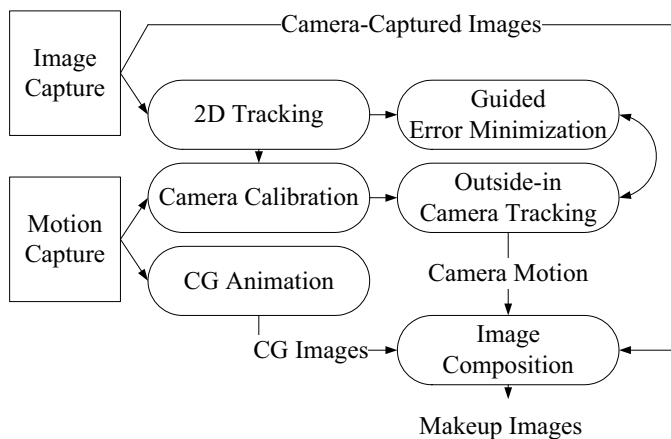


Figure 3. Data flow of our digital special makeup method

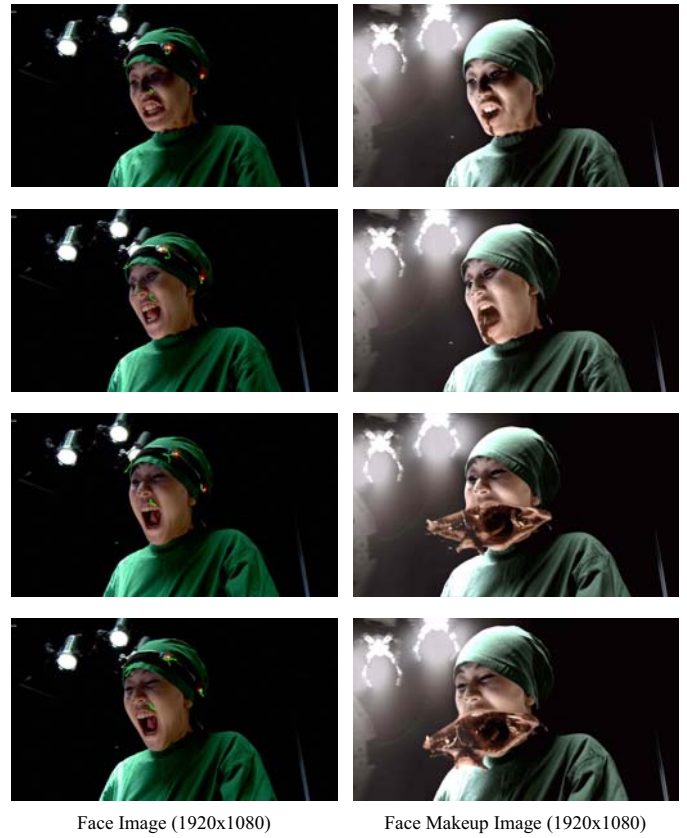


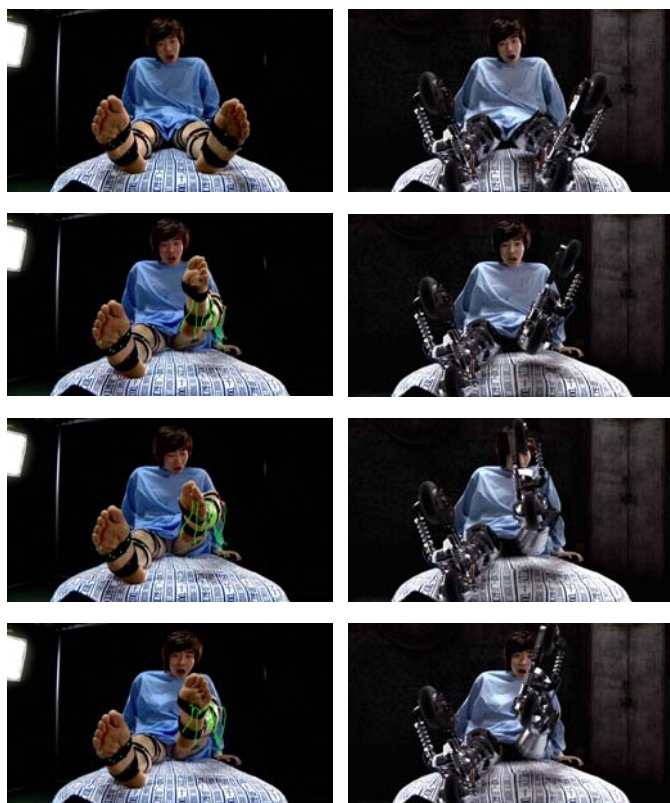
Figure 4. Result of face special makeup

IV. EXPERIMENTAL RESULTS

We verified the performance of our digital special makeup system with real motion capture and the video data. For experiment, we set the Impulse system to run in 120 fps. Eight active LED markers were attached to the camera target. Figures 4, 5, and 6 show the results of the digital special makeup for human face ($N=4$), leg ($N=16$), and body ($N=32$). The face and the leg images were captured in an indoor motion capture studio, but the body images in an outdoor environment. Note that our method allowed us to perform digital special makeup in outdoors as well as indoors. The performance of our outside-in camera tracking method was analysed with additional three datasets. The RMSEs for all datasets were listed in Table 1.

Table 1. RESULT OF PERFORMNACE ANALYSIS

Dataset	1	2	3
Number of frames	749	750	810
Number of markers $N(K)$	13(5)	52(7)	21(5)
RMSE	0.57	0.56	0.54



Leg Image (1920x1080)

Leg Makeup Image (1920x1080)

Figure 5. Result of leg special makeup



Body Image (1920x1080)

Body Makeup Image (1920x1080)

Figure 6. Result of body special makeup

V. CONCLUSIONS

We presented a new digital special makeup method to compose camera-captured human face or body images with CG ones for visual effects. We showed that our method could effectively combine an active motion capture technique with an outside-in camera tracking technique to simultaneously track the 3D motion of both human and camera with high accuracy. In addition, we showed that our method could be applied to digital special makeup in outdoors as well as indoors. To demonstrate the performance of our special makeup method, we presented various real digital special makeup results for human face, leg, and body.

ACKNOWLEDGMENT

This research was supported by the Ministry of Culture, Sports and Tourism (MCST) and the Korea Creative Content Agency (KOCCA) in the Culture Technology (CT) Research & Development Program 2010 [2-10-7602-003-10741-00-003].

REFERENCES

- [1] G. Welch and E. Foxlin, "Motion tracking: no silver bullet, but a respectable arsenal," *IEEE Computer Graphics and Applications*, vol. 22, no. 6, pp. 24-38, Nov.-Dec., 2002
- [2] C. Bregler, "Motion capture technology for entertainment," *IEEE Signal Processing Magazine*, vol. 24, no. 6, pp. 160-158, Nov. 2007
- [3] G. Lee, U. Yang, W. Son, Y. Kim, and D. Jo, "Virtual reality content-based training for spray painting tasks in the shipbuilding industry," *ETRI Journal*, vol. 32, no. 5, pp. 695-703, Oct. 2010
- [4] P. Havaldar, "Performance-driven facial animation," *ACM SIGGRAPH'06 Courses*, 2006
- [5] L. Sigal and A. O. Balan, "HumanEva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion," *International Journal of Computer Vision*, vol. 87, no. 1, pp. 4-27, Mar. 2010
- [6] J.-S. Kim, H.-M. Kim, and J.-H. Kim, "MoCap-based camera tracking for VFX," in *Proc. ICACT'10*, pp. 1614-1617, Feb. 2010
- [7] PhaseSpace Impluse System, <http://www.phasespace.com>
- [8] B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," *Journal of the Optical Society of America A*, vol. 4, no. 4, pp. 629-642, Apr. 1987
- [9] B. K. P. Horn, H. M. Hilden, and S. Negahdaripour, "Closed-form solution of absolute orientation using orthonormal matrices," *Journal of the Optical Society of America A*, vol. 5, no. 7, pp. 1127-1135, Jul. 1988
- [10] Weng, J., Cohen, P., and Herniou, M., "Camera Calibration with Distortion Models and Accuracy Evaluation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14(10): 965-980, 1992.
- [11] R. Hartley, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2ed., 2004
- [12] Antoniou, A. and Lu, W.-S., *Practical Optimization: Algorithms and Engineering Applications*, Springer, 2007.