A MATCH MOVING TECHNIQUE FOR MERGING CG AND HUMAN VIDEO SEQUENCES

Hirofumi SAITO† *Jun'ichi HOSHINO* ‡

† Niigata University, ‡ University of Tsukuba/PRESTO, JST

E-mail: hoshino@computer.org

ABSTRACT

Merging virtual object with human video sequence is an important technique for many applications such as special effects in movies and augmented reality. In a traditional method, the operator manually fit 3D body model onto the human video sequence, and generate virtual objects at the current 3D body pose. However, the manual fitting is a time consuming task, and the automatic registration is required. In this paper, we propose a new method for merging virtual objects onto the human video sequence. First, we track the current 3D pose of human figure by using the spatio-temporal analysis and the structural knowledge of human body. Then we generate CG objects and merge it with the human figure in video. In this paper, we demonstrate examples of merging virtual cloth with the video captured images.

1. INTRODUCTION

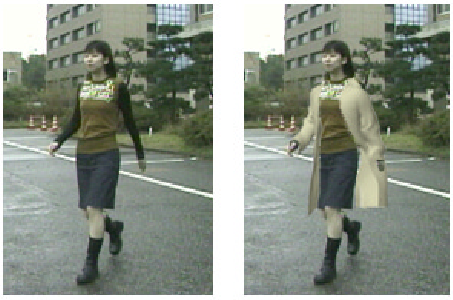
Merging computer generated objects onto human video sequence is an important technique for many applications such as special effects in movies and augmented reality applications. In a traditional method, the operator manually fit 3D body model onto the human video sequence, and generate virtual objects at the current 3D pose. However, manual fitting is a time consuming task, and automatic registration is required.

The automatic registration of virtual objects with video image is also called “match moving”[10]. The conventional techniques use the estimation of the camera position relative to the scene. The virtual objects are generated using the estimated camera position and orientation, and merged with the input image. The similar techniques are also investigated for the augmented reality applications [1,2,3,4,5,6,7]. However, merging virtual objects with the a complex, articulated figures such as a human body is still a difficult problem.

In this paper, we propose a new technique for the automatic registration of virtual objects with the human body images. As a typical example, we merge CG cloth onto the human video sequence. First, we track current 3D pose of human figure by using the spatio-temporal analysis and the structural knowledge of human body. Then we generate CG cloth and merge it with the human in video.

2. OVERVIEW

Figure 2 shows the overview of the algorithm. First, we estimate 3D pose data of human from video. We use 3D human model in Figure 4 to estimate the pose parameters of each parts. We represent a human body by an articulated structure consisting of 10 rigid parts corresponding body, head, upper arms (*ur-arm, ul-arm*), under arms (*ul-arm, ur-arm*), upper legs (*ur-leg, ul-leg*), under legs (*lr-leg, ll-leg*). The motion parameters are estimated using the spatial and temporal gradient method. The pose of human body at each frame is obtained from integration of a sequence of pose parameters onto the pose at the initial frame.



(a) original image (b) merging CG cloth

Figure 1: Example of merging CG cloth with a human image.

Then we generate CG objects using the pose parameters of the articulated objects. The self occlusion due to the body parts are also estimated using the pose parameters and the 3D shape of the body. Figure 1 shows the example of the merging virtual cloth onto the human video sequence. A possible application of this technique may be the virtual fashion simulator in

3. TRACKING HUMAN BODY

3.1 Human motion model

The body models are approximated by a polyhedron made by a CAD modeler. The model has a tree structure with a root at the trunk, and have a local coordinate system whose origin is located at a joint as described in Fig. 3. Rigid motion in each part of the body is a combination of rotation and translation. When a 3D point on the part j moves from $P_{oi}=(x_{oi},y_{oi},z_{oi},1)^T$ to P'_{oi}, the position is related by using the rigid body transform

$$P'_{oi} = E_i P_{oi}\ , \qquad E_i = \begin{bmatrix} Q_i & T_i \\ 0^t & 1 \end{bmatrix} \qquad (1)$$

where Q_i is the rotation matrix and T_i is the translation ma-

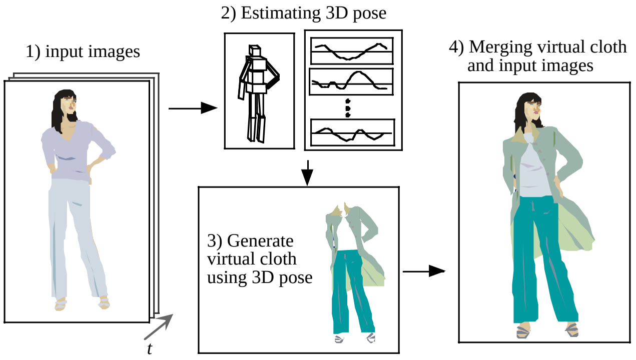


Figure 2: Overview of the algorithm

trix. Suppose that a point on *body1* moved to a new location, it can be calculated as follows.

1) Transform a point on the *body1* in the camera coordinate system P_s to the world coodinate system

$$P_w = E_s P_s \tag{2}$$

2) To move *body1*, we first transform a point in the world coordinate system P_w into the body coordinate system. Then apply a rigid body transform E_i to *body1*, and transform back to the world coordinate system. This operation can be written as $F_i = E_{oi}\,E_i\,E_{oi}^{-1}$.

$$P_w' = F_i P_w = F_i E_s P_s \tag{3}$$

3) Finally, transform a point P_w' into the camera coordinate sytem.

$$P_s' = E_s^{-1} P_w' = E_s^{-1} F_i E_s P_s \tag{4}$$

This transformation rule can be easily extended into a multi-body case.

$$P_{sj}^n = E_{sj}^{-1} F_1 \left[\prod_{m=0}^{n-2} F_{i-m} \right] E_{sj} P_s \tag{5}$$

Jacobian matrix can be derived from eq.(6) which will be used for the spatio-temporal analysis.

$$\begin{aligned} \frac{dP_s'}{dt} &= \sum_{j=1}^{2} \left\{ \frac{\partial p_i'}{\partial T_{xj}} \dot{T}_{xj} + \frac{\partial p_i'}{\partial T_{yj}} \dot{T}_{yj} + \frac{\partial p_i'}{\partial T_{zj}} \dot{T}_{zj} + \frac{\partial p_i'}{\partial \theta_{xj}} \dot{\theta}_{xj} + \frac{\partial p_i'}{\partial \theta_{yj}} \dot{\theta}_{yj} + \frac{\partial p_i'}{\partial \theta_{zj}} \dot{\theta}_{zj} \right\} \\ &= J(\phi)\dot{\phi} \end{aligned} \tag{6}$$

3.2 Estimation of motion parameters

The motion parameters are estimated using the spatial and temporal gradient method. Optical flow constraints in three dimensions can be written as

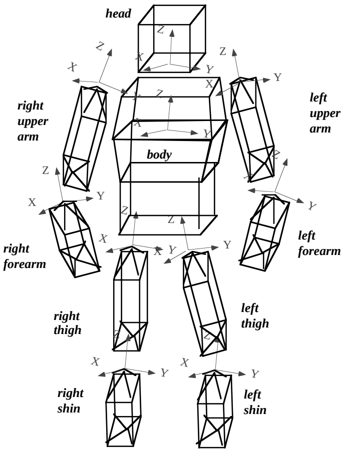


Figure 3: Human body model

$$G\dot{p} = -E_t \tag{7}$$

$$G = \left(f\frac{E_X}{z},\ f\frac{E_Y}{z}, \frac{X\,E_X\ + Y\,E_Y}{z} \right)$$

By substituting eq.(5) to eq(6), the motion parameters can be obtained using the least square method.

$$G^T J \dot{\phi} \ = -E_t \tag{8}$$

Eq. (7) can be extended to the multi camera form. Supposing the number of the cameras to be *n*, we have the *n* systems of linear equations which correspond to cameras. We can get one system of linear equations by gathering these *n* systems.

$$\begin{pmatrix} G_1^T J_1 \\ G_2^T J_2 \\ \vdots \\ G_n^T J_n \end{pmatrix} \dot{\phi} = \begin{pmatrix} -E_{t_1} \\ -E_{t_2} \\ \vdots \\ -E_{t_n} \end{pmatrix} \tag{9}$$

4. INITIAL REGISTRATION

The motion estimation technique in Sec. 3 only estimates inter-frame displacement of human body. Therefore we need an initial registration of body model to the input image. We estimate initial body pose by extracting the center line of each parts from 2D input image.

4.1 Modeling registration errors

Figure 4 shows the relationship of 2D center line on the image plane and a corresponding 3D body part. The 2D center line can be obtained using the silhouette of human body image and the principle axis calculation. The position and orientation of a body parts can be represented as $x=[r,t]^T$ where *t* is the translation and *r* is the rotation along the *xyz* axis. The error of a 2D center line and a 3D body part can be calculated as the minimum distance between 3D line segment *P=(p₁,p₂)* and plane *M*.

$$h(x, l)=\begin{bmatrix} (Rp_1 + t)\cdot n \\ (Rp_2 + t)\cdot n \end{bmatrix} \quad (10)$$

R is a3x3 matrix derived from *r*. *n* is a unit normal vector of plane *M*.

$$n=\frac{q_1' \times q_2'}{\left\| q_1' \times q_2' \right\|} \quad (11)$$

h(x,l) can be linearized using the Taylor expansion around the initial estimate $l = \hat{l}_i$ and $x = \hat{x}_{i-1}$.

$$h(x,l_i) \approx h(\hat{x}_{i-1},\hat{l}_i) + \frac{\partial h(\hat{x}_{i-1},\hat{l}_{ii})}{\partial x}(x - \hat{x}_{i-1}) + \frac{\partial h(\hat{x}_{i-1},\hat{l}_{ii})}{\partial l}(l - \hat{l}_i) \quad (12)$$

where $\frac{\partial h}{\partial x}, \frac{\partial h}{\partial l}$ are partial derivatives.

4.2 Minimizing registration errors

We minimize the errors between 2D center line and a 3D body parts by using the Extended Kalman Filter (EKF) [12]. The advantage of using EKF is that the estimation can be updated when the new measurement are available. The measurement equation is

$$z_i = H_i x + v_i$$

where

$$z_i = \frac{\partial h(\hat{x}_{i-1},\hat{l}_i)}{\partial x}\hat{x}_{i-1} - h(\hat{x}_{i-1},\hat{l}_i)$$

$$H_i = \frac{\partial h(\hat{x}_{i-1},\hat{l}_i)}{\partial x}, \qquad v_i = \frac{\partial h(\hat{x}_{i-1},\hat{l}_i)}{\partial l}(l_i - \hat{l}_i)$$

The optimization can be done using the Kalman Filter equations.

$$\hat{x}_i = \hat{x}_{i-1} + K_i[z_i - H_i\hat{x}_{i-1}]$$

$$K_i = \sigma_{i-1}H_i(H_i\sigma_{i-1}H_i^T + R)^{-1}$$

$$\sigma_i = \sigma_{i-1} - K_i H_i \sigma_{i-1}$$

where *K* is a Kalman Gain and σ_i is a covariance matrix of measurement. *R* is a covariance matrix of the estimates.

5. MERGING GRAPHIC OBJECTS

5.1 Cloth simulation

We generate Cloth CG using the pose parameters estimated in Sec.3 and Sec.4. We use MAYA Cloth™ for cloth simulation. Any other cloth simulation technique can be used for this purpose. We use the a approximate body model with the smooth surface to simulate natural shape of cloth CG. The size of the each parts are equal to the 3D body model used for the tracking.

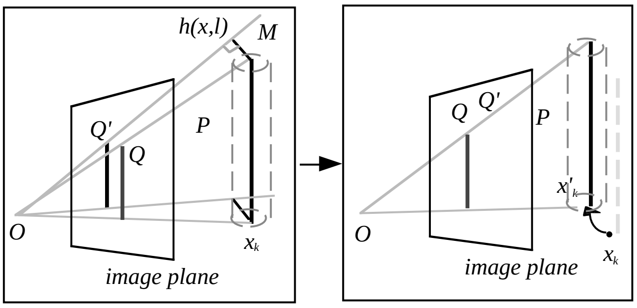


Figure 4: A relationship of the 2D center line and a corresponding body part

5.2 Dealing with occlusions

We can not simply overlay the virtual objects because it may be partially occluded by body parts. For example, user's hands may be before the virtual objects. We need relative depth information to calculate which body parts may occlude virtual objects. Again we use the approximate human body model to deal with occlusions.

1) First, we create the texture mapped 3D model of the human body. Because the body model is already registrated onto the input image, the intensity value of input image can be stored as a texture of 3D models.
2) The CG cloth and hairstyle are added to the texture mapped 3D model of human body.
3) Render the texture mapped body model and CG objects together to obtain synthesized image.

In our experiments, the small shape difference of the body model is not critical because we render from the same view point with the input image.

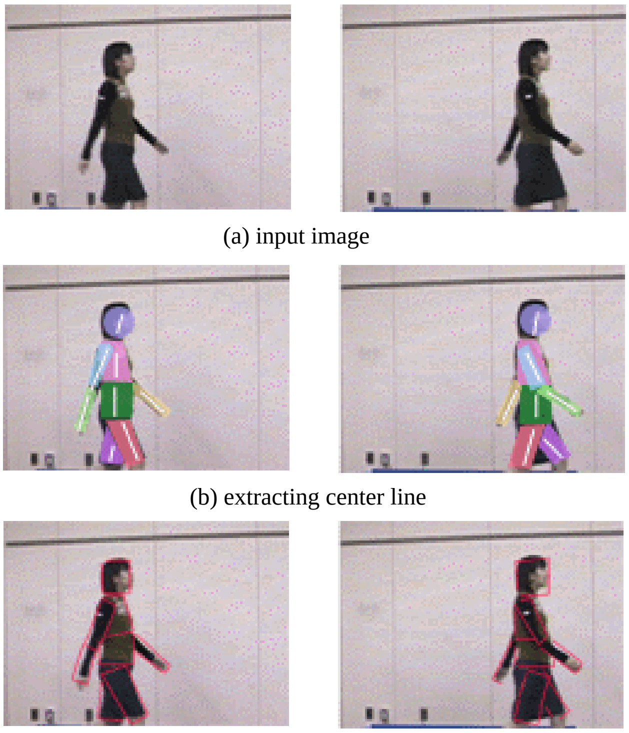
6. EXPERIMENTS

We have implemented a prototype system using PC to demonstrate the algorithm. Figure 5 shows the example of the initial model fitting. (a) is a input image, and (b) is the extracted center lines. (c) is the result of model fitting.

Figure 6 shows the movie sequence of a walking person. The input image is 50 frames of 640*480. (a) is 30 and 40 frame of the input image sequences. (b) is the tracking result. The wireframe of the human body model is overlaid onto the input images. (c) is the generated cloth image using the 3D body pose estimated from the input images. (d) is the input images with CG cloth.

7. CONCLUSION

In this paper, we proposed a new match moving technique for an articulated figures such as human. First, we track current 3D pose of human figure by using the spatio-temporal analysis and the structural knowledge of human body. Then we generate CG cloth and merge it with the human figure in video. We

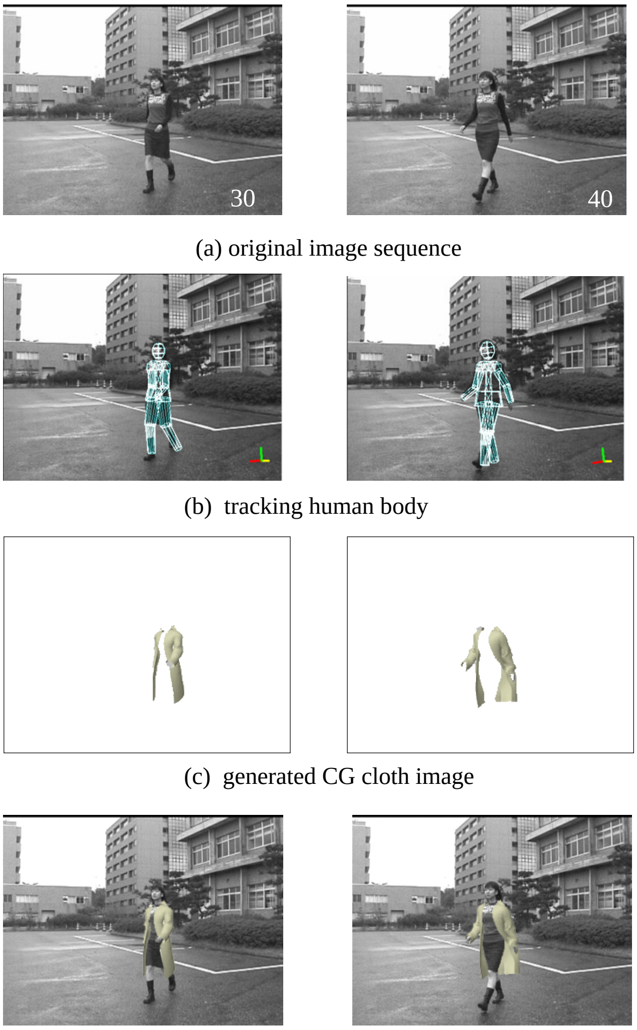


(a) input image

(b) extracting center line

(c) registration of 3D model

Figure 5: Result of automatic registration



(a) original image sequence

(b) tracking human body

(c) generated CG cloth image

(d) merging CG cloth with the input images sequence

Figure 6: Exmaple of merging virtual cloth

have demonstrated the examples using the video sequences. The future work includes the real-time implementation of the proposed technique.

8. REFERENCES

[1] V. Blanz, T. Vetter: “A Morphable Model for the Synthsis of 3D faces”, Proc. SIGGRAPH’99, pp.187-194, 1999

[2] M.Bajura, H.Fuchs and R.Ohbuchi: ``Merging virtual objects with the real world: Seeing ultrasound imagery within the patient", SIGGRAPH '92, pp. 203-210

[3] M.Bajura and U.Neumann: ``Dynamic Registration Correction in Video-Based Augmented Reality Systems", IEEE Computer Graphics and Applications, Vol.15, No.5, Sep. 1995, pp.52-60.

[4] M.Bajura and U.Neumann: ``Dynamic Registration Correction in Augmented Reality Systems", IEEE VRAIS 1995 Proc., 1995, pp.189-196.

[5] E.K.Edwards, J.P.Rolland and K.P.keller: ``Video See-Through Design for Merging of Real and Virtula Environments", IEEE VRAIS 1993 Proc., pp.222-233.

[6] S.Gottschalk and J.Hughes: ``Autocalibration for Virtual Environments Tracking Hardware", SIGGRAPH'93, pp.65-72.

[7] R.Azuma and G.Bishop: ``Improving Static and Dynamic Registration in an Optical See-Trough HMD", SIGGRAPH '94, pp.197-204.

[8] A.L.Janin, D.W.Mizwell and T.P.Caudell: ``Calibration of Head-Mounted Display for Augmented Reality Applications", IEEE VRAIS 1993 Proc., pp.246-255.

[9] D.M. Garvila:"Visual Analysis of Human Movement: A Survey", Computer Vision and Image Understanding, Vol.73, No.1, pp.82-98, 1999

[10] C. Barron:”Matte Painting in the Digital Age”, Animation Sketches, SIGGRAPH’98, 1998

[11] M. Yamamoto, A. Sato, S. Kawada,”Incremental tracking of human action from multiple views, Proc. of CVPR’98, pp.2-7, 1998

[12] H.W. Sorenson:"Kalman Filtering: Theory and Application", New York, IEEE Press, 1985