

# Scene Constraints-Aided Tracking of Human Body

Masanobu Yamamoto and Katsutoshi Yagishita  
Department of Information Engineering  
Niigata University  
8050, Ikarashi 2-nocho, Niigata 950-2181, Japan  
yamamoto@ie.niigata-u.ac.jp

## Abstract

*This paper describes a new method for tracking of a human body in 3D motion by using constraints imposed on the body from the scene. An image-based approach for tracking exclusively uses a geometrical model of the human body. Since the model usually has a large number of degrees of freedom (DOF), a chance to be corrupted by noise increases during the tracking process, and the tracking may fall in an ill-posed problem. To cope with this problem, we pay our attention to that a human body can not move freely, and usually receive some constraints from the scene. The new method uses constraints imposed on position, velocity and acceleration of the part of the body from the scene. These constraints can reduce the DOF of the model. This reduction guarantees the tracking problem to be a well-posed problem, and prevents tracking errors by noise. Experiments with real image sequences support a precise tracking of the body.*

## 1 Introduction

An image-based tracking of human bodies has many potential applications such as computer-human interface, motion capturing for animation character, surveillance for security, and so on. The image-based approach for tracking exclusively uses a geometrical model of the human body. Fitting the model to the body frame by frame [9, 10, 7, 3, 2, 6, 11, 4] results into a sequence of the pose of the body, i.e. tracking of the body. However, the frame by frame fitting needs a large amount of computational efforts.

A few researchers proposed an incremental approach as an alternative way [12, 13, 1]. This approach performs the tracking by accumulating pose increments between the successive frames onto the initial pose. Since a model fitting for once determines the initial pose, and the pose increment is estimated by linear calculation, the computational amount becomes smaller. The incremental approach has, however, an inherent drawback. It is that an accumulation of the estimation error falls in a failure of the tracking at last.

To cope with this problem, we pay our attention to that a human body can not move freely, and usually receive some constraints from the scene. We propose that a scheme of the incremental tracking should use the constraints from the scene.

Leung and Yang [7] proposed a use of scene constraints for tracking a gymnastic player. However, their constraint, namely “support”, is to fix a part of the body on a point in the scene. Our approach can constraint position, velocity and acceleration of the human body in motion.

The next section describes a motion model of human body. The Section 3 summarizes the image-based constraints. The Section 4 proposes several constraints from the scene. The Section 5 couples the scene constraints with the image-based constraints. The Section 6 shows that the coupled constraints result in successful tracking.

## 2 Motion Model

We represent a human body by an articulated structure model as shown in fig.1. This model consists of a set of the solid parts corresponding to parts of the body. The parts of the model are linked with each other in a hierarchy system. The pose of the model is given by angles between the linked parts and position of the top part in the hierarchy. Let a vector  $\mathbf{q} = (q_1, q_2, \dots, q_n)^T$  be a set of the pose parameters. Then, we can represent a relation between a position,  $\mathbf{p} = (x, y, z)^T$ , on the subject and the pose vector by the following function.

$$\mathbf{p} = \mathbf{f}(\mathbf{q}) \quad (1)$$

where an initial position is neglected.

The eq.1 is a nonlinear relation in general. However, a temporal derivation,  $\dot{\mathbf{q}}$ , of the pose,  $\mathbf{q}$ , can be linearly related with a velocity vector,  $\dot{\mathbf{p}}$ , using the Jacobian matrix  $\mathbf{J}(\mathbf{q})$ .

$$\dot{\mathbf{p}} = \mathbf{J}(\mathbf{q})\dot{\mathbf{q}} \quad (2)$$

The  $\dot{\mathbf{q}}$  denotes movements of the model. We call the  $\dot{\mathbf{q}}$  motion parameters.

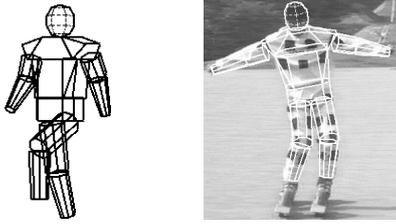


Figure 1: Fitting a body model to an image.

### 3 Motion Constraints from Images

Let an intensity on images in sequence be  $E(X, Y, t)$ , where  $(X, Y)$  and  $t$  denote a position on the image and time, respectively. A velocity vector,  $(U, V)$ , on the image is constrained by the equation [5],

$$E_X U + E_Y V + E_t = 0 \quad (3)$$

where  $E_X$ ,  $E_Y$  and  $E_t$  are gradients of the intensity in the directions of  $X$ ,  $Y$  and  $t$  of  $E_t$ , respectively.

Let us the constraint eq.(3) extend toward a constraint on the human movement beyond 3D movement in the scene.

We suppose the camera model to be a perspective projection with the focal length,  $f$ . The scene coordinates,  $(x, y, z)$ , is set so that the origin should be at the center of projection, the  $z$  axis the optical axis, and the image plane  $z = f$ . The origin on the image coordinates is put on an intersection between the image plane and the optical axis. The  $X$  and  $Y$  axes in the image plane are aligned along the  $x$  and  $y$  axes of the scene coordinate, respectively.

Then, a point  $\mathbf{p} = (x, y, z)^\top$  in the scene is projected onto a point  $(X, Y)$  in the image plane by the transformation

$$\begin{cases} X &= f \frac{x}{z} \\ Y &= f \frac{y}{z} \end{cases} \quad (4)$$

Differentiating eq.(4) with respect to time, we can get a relation between 3D velocity,  $\dot{\mathbf{p}} = (u, v, w)^\top$ , in the scene and 2D velocity,  $(U, V)$ , in the image such as

$$\begin{cases} U &= \frac{fu - Xw}{z} \\ V &= \frac{fv - Yw}{z} \end{cases} \quad (5)$$

Substituting eq.(5) into eq.(3), we have

$$\mathbf{G}^\top \dot{\mathbf{p}} = -E_t \quad (6)$$

where

$$\mathbf{G} = \left( f \frac{E_X}{z}, f \frac{E_Y}{z}, -\frac{XE_X + YE_Y}{z} \right)^\top \quad (7)$$

Moreover, eliminating  $\dot{\mathbf{p}}$  from eqs.(6) and (2), we finally obtain an image-based equation to constraint motion parameters,  $\dot{\mathbf{q}}$ .

$$\mathbf{G}^\top \mathbf{J} \dot{\mathbf{q}} = -E_t \quad (8)$$

Now, we explain a tracking process based on the constraint (8). At an initial frame, we fit the human model to the human body on the image plane. This model fitting, performed by hand-operated manner in this paper, may not so correct as model fitting of a rigid object (e.g. [8]) since we use a rough model of the human body and fit the model to an image from single camera view. However, even model fitting from single camera view can give a human pose which is not so far from the correct pose.

Motion parameters, pose increment between successive frames, can be obtained from a system of linear equations eq.(8) which correspond to feature points on the image of the human body.

$$\begin{pmatrix} \mathbf{G}_1^\top J_1 \\ \mathbf{G}_2^\top J_2 \\ \vdots \\ \mathbf{G}_n^\top J_n \end{pmatrix} \dot{\mathbf{q}} = \begin{pmatrix} -E_{t1} \\ -E_{t2} \\ \vdots \\ -E_{tn} \end{pmatrix} \quad (9)$$

where each suffix denotes a point on the human image.

Solving eq.(9) by a least-square method, we have the motion parameters,  $\dot{\mathbf{q}}$ . Tracking is performed by integration the motion parameters, pose increments, on the initial pose.

### 4 Motion Constraints from Scene

The subject usually moves under constraints in the scene. For example, when we are walking, a position of the foot is on the floor. Opening or closing the door, a hand is guided by a knob of the door. A train runs on the railroad. In skating and skiing, a velocity of the subject is parallel to the surface. When an object falls in the gravity field, the object is accelerated in the direction of the acceleration of gravity.

The scene imposes several constraints on a position, velocity and acceleration of a point of the object. We can expect more reliable tracking by adding these scene constraints to the image-based constraints described in the previous Section.

#### 4.1 Constraint on Position

We suppose that a point,  $\mathbf{p}$ , on the subject is constrained on a plane,

$$\mathbf{n}^\top \mathbf{x} = d \quad (10)$$

where  $\mathbf{n}$  is a normal vector, and  $d$  is a distance between the origin and the plane. When the point,  $\mathbf{p}$ , moves to some position,  $\mathbf{p}'$ , at the next frame, the movement,  $\Delta \mathbf{p}$ , satisfies an equation,

$$\mathbf{p}' = \mathbf{p} + \Delta \mathbf{p} \quad (11)$$

On the other hand, since the new position,  $\mathbf{p}'$ , also should be on the plane represented by eq.(10), we have

$$\mathbf{n}^\top(\mathbf{p} + \Delta\mathbf{p}) = d \quad (12)$$

We suppose that a frame interval is set at a unit, 1, and an apparent movement between successive frames is very small, that is,

$$\Delta\mathbf{p} = \dot{\mathbf{p}} \quad (13)$$

Rewriting eq.(12) by eq.(13), we have

$$\mathbf{n}^\top(\mathbf{p} + \dot{\mathbf{p}}) = d \quad (14)$$

Moreover, substituting eq.(2) into eq.(14), we can finally get ,

$$\mathbf{n}^\top \mathbf{J} \dot{\mathbf{q}} = d - \mathbf{n}^\top \mathbf{p} \quad (15)$$

This equation is a constraint of motion parameters,  $\dot{\mathbf{q}}$ , in order to keep holding a point of the subject onto the plane.

In case of constraining to the curved surface, we use a tangent plane of the cured surface. This is a natural extension of the plane constraint (15) from viewing that the surface normal,  $\mathbf{n}$ , is considered as a function of position,  $\mathbf{p}$ .

A curved line can be an intersection between two curved surface patches. Also, a point can be an intersection among three independent planes. Therefore, simultaneously constraining to the two cured surface patches results into constraining to the curved line. Similarly, constraining the three planes results into constraining the fixed point.

#### 4.2 Constraint on Velocity

A velocity vector,  $\dot{\mathbf{p}}$ , of a point,  $\mathbf{p}$ , on the subject supposes to be constrained in the direction,  $\mathbf{s} = (s_x, s_y, s_z)^\top$ . Since this means that  $\dot{\mathbf{p}}$  has the same direction as  $\mathbf{s}$ , we have

$$\mathbf{s} \times \dot{\mathbf{p}} = \mathbf{o} \quad (16)$$

Substituting eq.(2) into eq.(16), we have

$$\mathbf{s} \times (\mathbf{J} \dot{\mathbf{q}}) = \mathbf{o} \quad (17)$$

Moreover, expanding the vector product in eq.(17), we have

$$\begin{cases} (0, -s_z, s_y) \mathbf{J} \dot{\mathbf{q}} = 0 \\ (s_z, 0, -s_x) \mathbf{J} \dot{\mathbf{q}} = 0 \\ (-s_y, s_x, 0) \mathbf{J} \dot{\mathbf{q}} = 0 \end{cases} \quad (18)$$

These equations are also constraints of motion parameters,  $\dot{\mathbf{q}}$ , so that the velocity vector could be parallel to the specified vector,  $\mathbf{s}$ .

We have one more way to constraint the velocity vector. We suppose that a velocity vector  $\dot{\mathbf{p}}$ , of a point,

$\mathbf{p}$ , should be parallel to a plane with surface normal,  $\mathbf{n}$ , that is,

$$\mathbf{n}^\top \dot{\mathbf{p}} = 0 \quad (19)$$

Substituting the  $\dot{\mathbf{p}}$  of eq.(2) into (19), we have,

$$\mathbf{n}^\top \mathbf{J} \dot{\mathbf{q}} = 0 \quad (20)$$

This equation is a constraint of motion parameters,  $\dot{\mathbf{q}}$ , in order to keep the velocity vector to be parallel to the plane with surface normal,  $\mathbf{n}$ .

#### 4.3 Constraint on Acceleration

When an acceleration vector,  $\ddot{\mathbf{p}}$ , of a point,  $\mathbf{p}$ , on the subject has the same direction as a specified vector,  $\mathbf{s} = (s_x, s_y, s_z)^\top$ , as is analogous to the case of velocity constraints, we have

$$\mathbf{s} \times \ddot{\mathbf{p}} = \mathbf{o} \quad (21)$$

Supposing the acceleration to be almost constant in time, and the frame interval to be a unit, 1, the acceleration,  $\ddot{\mathbf{p}}$ , is obtained from subtraction between velocities,  $\dot{\mathbf{p}}$ , at successive frames.

$$\ddot{\mathbf{p}}_i = \dot{\mathbf{p}}_i - \dot{\mathbf{p}}_{i-1} \quad (22)$$

Substituting eq.(22) into eq.(21), we have

$$\mathbf{s} \times \dot{\mathbf{p}}_i = \mathbf{s} \times \dot{\mathbf{p}}_{i-1} \quad (23)$$

Moreover, substituting eq.(2) into eq.(23), we have an equation with respect to motion parameters. Since

$$\begin{aligned} \Delta\mathbf{p}_{i-1} &= -(\mathbf{J}_i(-\Delta\mathbf{q}_{i-1})) \\ &= \mathbf{J}_i \Delta\mathbf{q}_{i-1} \approx \mathbf{J}_i \dot{\mathbf{q}}_{i-1} \end{aligned} \quad (24)$$

, if we use the Jacobian matrix at the current frame  $i$ , we can finally get

$$\mathbf{s} \times \mathbf{J} \dot{\mathbf{q}}_i = \mathbf{s} \times \mathbf{J} \dot{\mathbf{q}}_{i-1} \quad (25)$$

Expanding the vector product in eq. (25), each component is;

$$\begin{cases} (0, -s_z, s_y) \mathbf{J} \dot{\mathbf{q}}_i = (0, -s_z, s_y) \mathbf{J} \dot{\mathbf{q}}_{i-1} \\ (s_z, 0, -s_x) \mathbf{J} \dot{\mathbf{q}}_i = (s_z, 0, -s_x) \mathbf{J} \dot{\mathbf{q}}_{i-1} \\ (-s_y, s_x, 0) \mathbf{J} \dot{\mathbf{q}}_i = (-s_y, s_x, 0) \mathbf{J} \dot{\mathbf{q}}_{i-1} \end{cases} \quad (26)$$

These equations are constraints,  $\dot{\mathbf{q}}_i$ , in order to keep the acceleration vector,  $\ddot{\mathbf{p}}$ , to be parallel to a specific vector,  $\mathbf{s}$ .

When the acceleration vector,  $\ddot{\mathbf{p}}$ , is kept to be parallel to a specific plane with a surface normal,  $\mathbf{n}$ , the corresponding constraint is

$$\mathbf{n}^\top \ddot{\mathbf{p}} = 0 \quad (27)$$

Again, using eqs.(22), (2) and (24), we can rewrite the eq.(27) as

$$\mathbf{n}^\top \mathbf{J} \dot{\mathbf{q}}_i = \mathbf{n}^\top \mathbf{J} \dot{\mathbf{q}}_{i-1} \quad (28)$$

This equation is a constraint of motion parameters,  $\dot{\mathbf{q}}_i$ , in order to keep the acceleration vector to be parallel to a specific plane with a surface normal,  $\mathbf{n}$ .

If it is necessary to constraint velocity and acceleration on a curved surface, we can use a tangent plane which approximates the curved surface.

## 5 Coupling Constraints

We described many kinds of constraints; eq.(8) in Section 3, and eqs.(15), (18), (20), (26), (28) in Section 4. Since all constraints are linear equations with respect to motion parameters,  $\dot{\mathbf{q}}$ , it is easy to couple all in one. We have a large system of linear equations,

$$\begin{pmatrix} w_1 \mathbf{G}_1^\top \mathbf{J}_1 \\ \vdots \\ w_j \mathbf{n}_j^\top \mathbf{J}_j \\ \vdots \\ w_k \hat{\mathbf{s}}_{kx}^\top \mathbf{J}_k \\ w_k \hat{\mathbf{s}}_{ky}^\top \mathbf{J}_k \\ w_k \hat{\mathbf{s}}_{kz}^\top \mathbf{J}_k \\ \vdots \\ w_l \mathbf{n}_l^\top \mathbf{J}_l \\ \vdots \\ w_m \hat{\mathbf{s}}_{mx}^\top \mathbf{J}_m \\ w_m \hat{\mathbf{s}}_{my}^\top \mathbf{J}_m \\ w_m \hat{\mathbf{s}}_{mz}^\top \mathbf{J}_m \\ \vdots \\ w_n \mathbf{n}_n^\top \mathbf{J}_n \\ \vdots \end{pmatrix} \dot{\mathbf{q}}_i = \begin{pmatrix} -w_1 E_{t1} \\ \vdots \\ w_j (d_j - \mathbf{n}_j^\top \mathbf{p}_j) \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ w_m \hat{\mathbf{s}}_{mx}^\top \mathbf{J}_m \dot{\mathbf{q}}_{i-1} \\ w_m \hat{\mathbf{s}}_{my}^\top \mathbf{J}_m \dot{\mathbf{q}}_{i-1} \\ w_m \hat{\mathbf{s}}_{mz}^\top \mathbf{J}_m \dot{\mathbf{q}}_{i-1} \\ \vdots \\ w_n \mathbf{n}_n^\top \mathbf{J}_n \dot{\mathbf{q}}_{i-1} \\ \vdots \end{pmatrix} \quad (29)$$

where

$$\begin{cases} \hat{\mathbf{s}}_{ox} = (0, -s_{oz}, s_{oy})^\top \\ \hat{\mathbf{s}}_{oy} = (s_{oz}, 0, -s_{ox})^\top \\ \hat{\mathbf{s}}_{oz} = (-s_{oy}, s_{ox}, 0)^\top \end{cases} \quad (30)$$

The matrix and the constant vector in eq.(29) consist of six blocks for the sake of convenience. The first block corresponds to the image-based constraint eq.(8). The remaining blocks represent the constraints from the scene properties. The blocks from the 2nd to sixth correspond to the position, eq.(15), velocity, eqs.(18), (20), and acceleration, eqs.(26), (28) constraints, respectively.

When the subject is constrained on a curved surface, the corresponding tangent plane with surface normal,

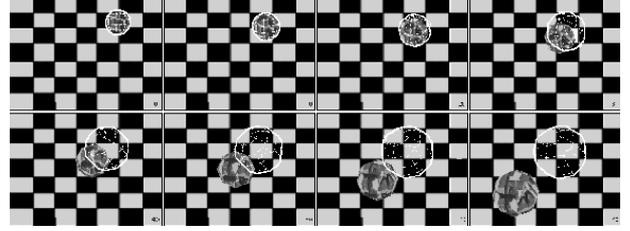


Figure 2: Tracking result by the only image-based method.

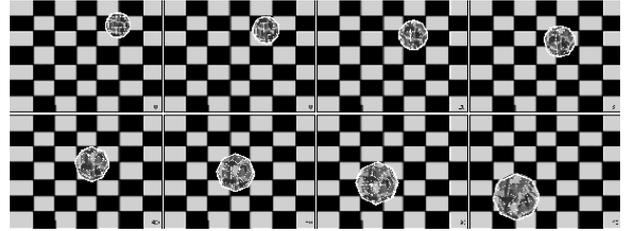


Figure 3: Tracking result by the image-based method with scene constraints.

$\mathbf{n}$ , and distance from the origin,  $d$ , are functions of time denoted by frame number  $i$ . In the case to be constrained on curved line or fixed point in the scene, 2 or 3 constraints are imposed on a point of the subject.

Solving the eq.(29) by least-square method, we can estimate motion parameters,  $\dot{\mathbf{q}}$ . This solution enables us to make tracking based on the image-based and scene constraints. The  $w$  in eq.(29) is a weighting imposed on each constraint. Harder weighting on the scene constraints rather than the image based constraints may lead to a solution without movement, since the eq.(29) becomes biased to the scene constraints. Otherwise, the eq.(29) is biased to the image based constraints. Therefore, we have to carefully choose the magnitude of each weight on the constraint.

## 6 Experiments of Tracking

We will show the tracking results based on the proposed approach. A simulation in the sub-section 6.1 will present a quantitative improvement by adding scene constraints. Tracking results in real image will give in the sub-sections 6.2 and 6.3.

### 6.1 Tracking in Synthesized Images

We tracked a ball rolling down on a slope. The number of images in the sequence is 70. A Gaussian noise corrupts each image. Position, and velocity and acceleration vectors of the center of gravity of the ball are constrained in parallel to the sloping surface.

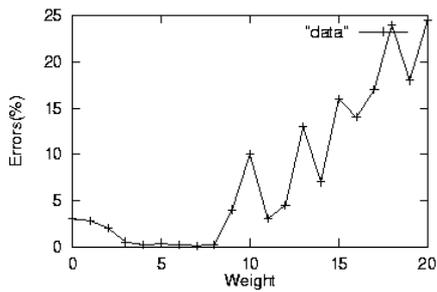


Figure 4: Tracking error versus the magnitude of weight of scene constraints.



Figure 5: Tracking result by the only image-based method.

Figs.2 and 3 show tracking results, where images are corrupted by the Gaussian noise with mean the value, 0, and the deviation, 10. If the scene constraints are not used, the tracking falls in failure because of the noise and strong texture in the background. The scene constraints can guarantee a successful tracking. A ratio of the error to the translation of center of the ball from the initial and final frames was 2.7%. A norm of the difference between the two matrixes, expressing an orientation of the ball, at initial and final frames was 0.12.

Fig.4 shows percentages of the error ratio of the translation at various weight, where we fixed the weight to image-based constraints at 1, and varied all the weight to the scene constraints coherently. This experiment shows that the harder weighting makes a correct tracking be difficult. We chose 4 for the weight in the experiment of tracking ball, and use it in the remaining experiments.

### 6.2 Tracking a Real Arm

We show a result of tracking an arm opening and closing a door during the 100 frames. A hand of the arm is constrained to the trace of a knob of the door. The trace is an intersection between a surface of the cylinder, of which axis is aligned with rotational axis of the door, and a horizontal plane at a height of the

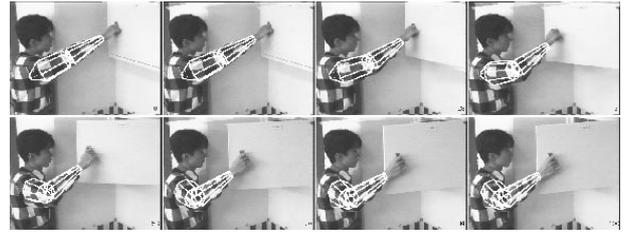


Figure 6: Tracking result by the image-based method with scene constraints.

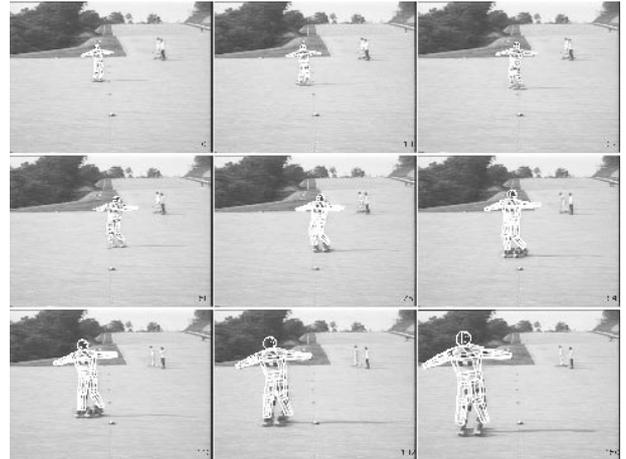


Figure 7: Tracking result by the only image-based method.

knob. The geometry of the cylinder and the horizontal plane is given by measurement of the real scene.

Fig.6 shows a tracking result, while fig.5 shows a tracking result without the scene constraints. The tracking result of the hand with scene constraints has little deviation, since the position of the hand is constrained by two surfaces, say, cylinder and plane. This successful tracking leads to a good tracking result of the whole arm too.

### 6.3 Tracking a Skier

We show a result of tracking a skier in slalom on the slope during 150 frames. The skier is constrained to the surface of the slope. We use the scene constraints imposed on position and velocity. The constraint on the position works so that the both feet of the skier could move on the slope. The constraint on the velocity works so that velocities at the both feet and a center of the skier could be parallel to the slope. Fig.8 shows tracking result, while fig.7 shows a tracking result using only image-based constraints without the scene constraints. Fig.9 denotes the side views of tracking re-

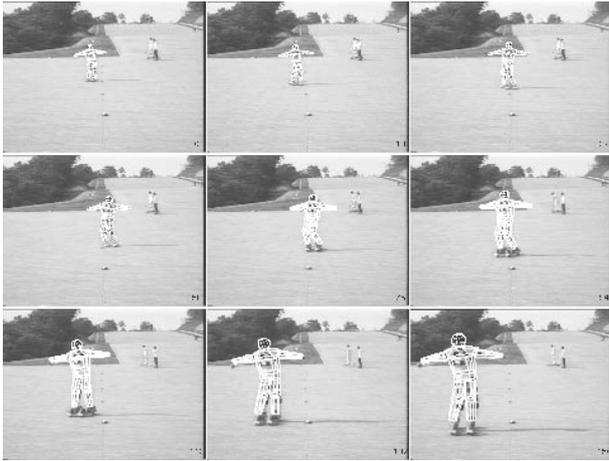


Figure 8: Tracking result by the image-based method with scene constraints.

sults for both cases at final frame. The tracking with scene constraints keeps the both feet on the slope. On the other hand, tracking without scene constraints fails in keeping one foot on the slope, and the pose does not look like one in skiing.

## 7 Conclusions

We proposed that tracking of human body should be performed under constraints from the scene. Coupling the scene constraints and the traditional image-based constraints makes the accuracy of tracking results improve through the experiments.

This method may be able to apply to the case where the 3D shape of the scene is easily given; a room under surveillance, and field and gymnasium for sports.

## References

- [1] C.Bregler and J.Malik, "Tracking people with twist and exponential maps", Proc. IEEE CVPR'98, pp.8-15, 1998.
- [2] D.M.Gavrila and L.S.Davis, "3-D model-based tracking of humans in action: a multi-view approach", Proc. IEEE CVPR'96, pp.73-80, 1996.
- [3] L.Goncalves, E.D.Bernardo, E.Ursella and P.Perona, "Monocular tracking of the human arm in 3D", Proc. 5th ICCV, pp.764-770, 1995.
- [4] I.Haritaoglu, D.Harwood and L.S.Davis, "W<sup>4</sup>: Who? When? Where? What? A real time system for detecting and tracking people, Proc. FG'98, pp.222-227, 1998.
- [5] B.K.P.Horn and B.G.Schunck, "Determining optical flow", Artificial Intelligence, **17**, 1-3, pp.185-203, 1981
- [6] I.A.Kakadiaris and D.Metaxas, "Model-based estimation of 3D human motion with occlusion based on active multi-viewpoint selection", Proc. IEEE CVPR'96, pp.81-87, 1996.
- [7] M.K.Leung and Y.H.Yang, "First sight: A human body outline labeling system", IEEE Trans.PAMI, **17**, 4, pp.359-377, 1995
- [8] D.G.Lowe, "Fitting Parameterized Three-dimensional Models to Images", IEEE Trans.PAMI, **13**, 5, pp.441-450, 1991
- [9] J.M.Rehg and T.Kanade, "DigitEyes: Vision-based human hand tracking", CMU-CS-93-220, 1993.
- [10] K.Rohr, "Towards model-based recognition of human movements in image sequences", CVGIP: Image Understanding, **59**, 1, pp.94-115, 1994
- [11] C.Wren, A.Azarbayejani, T.Darrell and A.Pentrand, "Pfinder: real-time tracking of the human body, IEEE Trans.PAMI, **19**, 7, pp.780-785, 1997
- [12] M.Yamamoto and K.Koshikawa, "Human motion analysis based on a robot arm model", Proc. IEEE CVPR'91, pp.664-665, 1991
- [13] M.Yamamoto, A.Sato, S.Kawada, T.Kondo and Y.Osaki, "Incremental tracking of human actions from multiple views", Proc. IEEE CVPR'98, pp.2-7, 1998

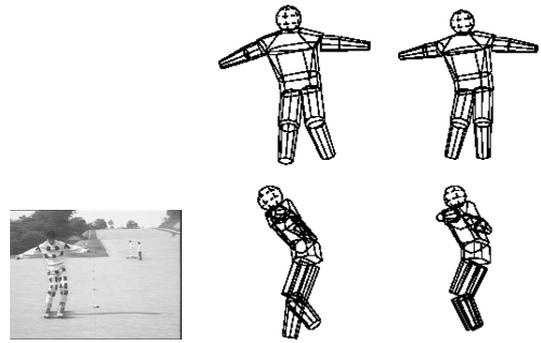


Figure 9: Left is the final frame. Center: 3D pose at the final frame by the only image-based method. Right: 3D pose at the final frame by the image-based method with scene constraints.