MVA '92 IAPR Workshop on Machine Vision Applications Dec. 7-9,1992, Tokyo

COMBINING MOTION SMOOTHNESS AND GREYSCALE CONSISTENCY IN ANALYSING REAL WORLD HUMAN BODY MOTION

Mo Weiguo
Department of Computer Science,
Fudan University,

Shanghai, 200433
P.R.China

ABSTRACT

This paper presents a technique combining MOTION SMOOTHNESS and GREYSCALE CONSISTENCY to analyse a long sequence monocular image of real world human body motion. Body joint point is chosen as feature point in several first frames manually, and the corresponding point in following frames is decided by computer through PREDICTING and MATCHING steps. Detail experiment on real images taken in complex background shows this method is qualified to be used in real circumstances.

Ⅰ. INTRODUCTION

There are several fields where the analysis of body motion image sequences plays an important role. Posture and motion analyses in athletic sports and in the training of physicially disabled persons are good examples.The traditional methods of motion analysis go as follows. The most primitive one is to manually indicate the feature points (usually body joint points, which can depict postures precisely according to anatomy and sports biomechanics) in each frame of a sequence of monocular temporal images. Their coordinates are integrated by machine to caculate their speeds and trajectories. Another method[1] is to put some marks on those places to which we pay particular attention prior to the motion.The analysis of the motion is done by detecting these marks in each image frame.

In recent years, several advanced methods have been proposed. Koichiro Akita[2] adopts a method of segment and discription under the conduction of human body model based on rigid body assumption. Maylor K. Leung[4] disusses a new technique for partitioning a human body in motion into meaningful parts. And I.K. Sethi[5] does an experiment which directly extrcats feature point from consective frames and matches these feature points using relaxation algorithm. The effects of all of these works mostly relies on the success of some low level processes,which usually quite difficult in real world circumstance. On the other hand, Lu Qin [3] proposes a method based on grey-scale similarity. Though this method is not very stable in some cases with complex background, it is simple, derective and reasonably based. Besides, we find its unstability is due to loose constrains. Thus we importing the concept of motion smoothness[8], and propose a "hybrid" method based on [3].

Ⅱ. ALGORITHM DISCRIPTION

Our task is: with a long sequence images having very small constant interval between two consective frames as input,we need to get the coordinates of some joint points(they are chosed manually in several first frames)in every frame as output, and thus trajectories of each feature point can be formed.

Because of the samll interview, we suppose:
MOTION SMOOTHNESS:[7][8]

(1)The scalar velocity, or speed of a given point,is relatively unchanged from one frame to the next; (2) The direction of motion of a given point is relatively unchanged from one frame to the next; (3)The uncertain ty of the motion direction of a given point increases as its speed decreases;
GREY-SCALE CONSISTENCY:[3][6]

(4)The grey-scale distribution within a certain limited area enclosing a given point is relatively unchanged from one frame to the next.

Accoording to these constrains, our task can be devided into two subroutines:
(1)Predicting Step:

(a) for every selected points a_K in frame K, a search region in frame K+1 is predicted from the information provided by a_{K-2}, a_{K-1} , a_K on the basis of assumption 1—3. The region varies according to the change of acceleration and velocity, which is actually a function of a_{K-2} ,a_{K-1} and a_K. (b)every point in this region is compared with a_K using grey-scale similarity[9][10]. Only those points with similarity great than a threshold T got further consideration. From these points the N most "similar" points are selected as candidate points of the next matching step.N varies with the similarity and isdecided by an adaptive equation: high the similarity,smaller the N,and vice versa.
(2)Matching Step:

Using an evaluation function based on assumption 1—4, an evalution value is got for every point in the

N-set. The point with the greatest value is choosen as the corresponding point.

The whole algorithm is consist of following four steps:

A. Search Region Predictation

Let t be the constant interval between two consective frames, P_k be feature point in Frame K, V_k be average velocity $V_k=(P_{k+1}-P_k)/t$, A_k be average acceleration $A_k=(V_{k+1}-V_k)/t$. We predict two P'_{k+1} and P''_{k+1} from privious velocity and acceleration respectively as follows:

$$P'_{k+1}=P_k+V_k*t=P_k+V_{k-1}*t=2P_k-P_{k-1}$$
($V_k=V_{k-1}$, uniform velocity motion)

$$P''_{k+1}=P_k+V_k*t=P_k+(V_{k-1}+A_{k-1}*t)*t=P_k+(V_{k-1}+A_{k-2}*t)*t$$
$$=P_k+(V_{k-1}+V_{k-1}-V_{k-2})*t=P_k+(2P_k-3P_{k-1}+P_{k-2})*t$$
$$=3P_k-3P_{k-1}+P_{k-2}$$
($A_{k-1}=A_{k-2}$, uniform acceleration motion)

Having settled P'_{k+1} and P''_{k+1}, connect them to get the middle point M, and then connect P_k and M. Let P_kM be the axis of the fan-type search region. The vertex angle β is caculated as [8], which is assumed to be inversly proportional to the variable V and is a function of the variable V. The function is decided by V_{max}, β_{max} and β_{min}. Having decided the vertex angle and the axis, the fan-type search region has been predicted.

B. Grey-Scale Similarity Caculation

Suppose P'_k is the ith candidate point in frame K+1. A window of dimwnsion (2u+1)*(2v+1) is centered on the point P'_k and it is correlated with a same sized window in the precede framecentered on P_k accordig to the following equation:

We choose this cross-correlation because it is in sensitive to absolute brightness and contrast in the two windows, and is useful if the two frames were obtained underdifferent illuminations.

Only those points with similarity great than threshold T(we set T=0.7) are selected and from these selected points, the top N points with max similarity consist the set of candidate points. Usually when the two frames are highly matohrd and the Φ value is very high, the predicted points seem to be "right" ones and are more reliable. In this case, the truely corres ponding point is more likely among theseveral ones with top Φ value. Thus a small number of candidate points are needed to do corresponding caculation. Otherwise more points are needed to ensure the truely corresponding one is among them. Thus N is inversly proportional to similarity value. Here we define an adaptive equation to choose N which is proved to be quite effective.

$$N = 1+\left[\frac{\theta}{\Phi_{max}^4}\right]$$

$\Phi_{max}=\max(\Phi'_{k,k+1})$. θ is a constant and $\theta<1$, $\theta\to 1$(we set $\theta=0.99$). [x] means integer nearest to but smaller than x. Suppose $\Phi_{max}=1$, then N=1, which means we only have one cadidate and it is the corresponding one; if $\Phi_{max}=0.7$, then N=5, sowe have 5 cadidates and this is also the maximum number of points in N-set.

C. Correspondence Problem

In this step N candidate points are as input. To evaluate the probobility of every point as correcponding one, we define an evaluation function considerating both motion and grey-scale consistency.

Let $P_{k-1}P_k$ denote the vector and $\|P_{k-1}P_k\|$ denote thedistance between P_{k-1} and P_k, then

if $\|P_{k-1}P_k\|>0$ and $\|P_kP_{k+1}\|>0$,
$F_e(K,K+1)=A+B+C=W_1*(P_{k-1}P_k \cdot P_kP_{k+1})/(\|P_{k-1}P_k\|*\|P_kP_{k+1}\|)+W_2*2(\|P_{k-1}P_k\|*\|P_kP_{k+1}\|)^{1/2}/(\|P_{k-1}P_k\|+\|P_kP_{k+1}\|)+W_3*\Phi_{k,k+1}$
if $\|P_kP_{k+1}\|>0$ and $\|P_{k-1}P_k\|=0$,
$F_e(K,K+1)=(W_1+W_2)*(1-\|P_kP_{k+1}\|/V_0)+W_3*\Phi_{k,k+1}$;
if $\|P_kP_{k+1}\|=0$ and $\|P_{k-1}P_k\|>0$,
$F_e(K,K+1)=W_2*(1-\|P_{k-1}P_k\|/V_0)+(W_1+W_3)*\Phi_{k,k+1}$;
if $\|P_kP_{k+1}\|=\|P_{k-1}P_k\|=0$,
$F_e(K,K+1)=(W_1+W_2+W_3)*\Phi_{k,k+1}$.

$$\Phi'_{k,k+1}=\sum_{\zeta=i-u}^{i+u}\sum_{\eta=j-v}^{j+v}\{F_k(\zeta,\eta)F_{k+1}(\zeta',\eta')-\mu_k(i,j)\mu_{k+1}(i',j')\}*\left\{\frac{\sigma_k(i,j)\,\sigma_{k+1}(i',j')}{(2u+1)(2v+1)}\right\}^{-1};$$

$\zeta'=\zeta+\|P'_{k+1}P_k\|_x$, $\eta'=\eta+\|P'_{k+1}P_k\|_y$, $i'=i+\|P'_{k+1}P_k\|_x$, $j'=j+\|P'_{k+1}P_k\|_y$;

where

$$\mu_k(i,j)=\frac{1}{(2u+1)(2v+1)}\sum_{\zeta=i-u}^{i+u}\sum_{\eta=j-v}^{j+v}F_k(\zeta,\eta),\quad \mu_{k+1}(i',j')=\frac{1}{(2u+1)(2v+1)}\sum_{\zeta=i-u}^{i+u}\sum_{\eta=j-v}^{j+v}F_{k+1}(\zeta',\eta');$$

are the window means for frame K and K+1 respectively;

$$\sigma_k^2=\frac{1}{(2u+1)(2v+1)}\sum_{\zeta=i-u}^{i+u}\sum_{\eta=j-v}^{j+v}\{[F_k(\zeta,\eta)]^2-[\mu_k(i,j)]^2\},$$
$$\sigma_{k+1}^2=\frac{1}{(2u+1)(2v+1)}\sum_{\zeta=i-u}^{i+u}\sum_{\eta=j-v}^{j+v}\{[F_{k+1}(\zeta',\eta')]^2-[\mu_{k+1}(i',j')]^2\};$$

are the window variances for frame K and K+1 respectively.

This function has value in [0,1]. A,B,C denote directional, speed, grey-scale term with weight W_1, W_2, W_3 respectively. $W_1+W_2+W_3=1$. The great the speed, direction and grey-scale deviation, the smaller the evaluation value. The point having thegreatest evaluation value is chosen as the true point on thetrajectory.

D.Correction And Human Interfare

In step B,when no candidate point with similarity great than T is obtained, we first consider the possibility of wrong assignment of a_k to the trajectory due to its closeness to the right one. In this case we select point one by one from the N-set of frame K according to its distance from $a_k$as a replace of the old a_k and caculate the new $\Phi'_{k,k+1}$ again untilget the usual N-set of frame K+1. But in a very few cases,this correction procedure also fails, then we suppose that the grey-scale changes abruptly. In this case human interfare isimposed. The P'_{k+1} and P''_{k+1} are proposed as reference points.We can choose one out of these two or an other one if we think it is more proper in the search region as the corresponding point. Once it has been selected, smoothness method similiar to that iterative optimization algorithm in [7] is employed.

Ⅳ. EXPERIMENT RESULTS AND CONCLUSION

We use a high speed television camera,which takes 100 frames per second. Suppose a sequence images of 100m dash is taken in real worldcircumstances. The ave rage running speed is 10 metres per second, thus the disparity between two consective frames is d=0.1m. The distance from camera to object is f_d=40—50m. The focal length is f_1=100mm and the film size is l_x=36mm by l_y=24mm, which is digitized to be s_x=512 pixel by s_y=512 pixel image. Thusthe 0.1m disparity in image(X-axis) would be (see Fig 2)

$$x = d*f_1/f_d = 0.1*100/50 = 0.2 \text{ mm}$$
$$= 0.2*s_x/l_x = 0.2*512/36 = 3 \sim 4 \text{ pixel}$$

which means the max change in one direction between two frames is less than 4 pixels, which we think is small enough to regard as being consistant and usually abrupt change of greyscale distribution will not occur.

Since postures of human body is usually described by 7 pairs,13 joints, and many human body movements, such as walking and running, are symmetry, we only choose the joints on the visible side as feature point. Thus we get seven feature points: A.head; B.should; C.elbow; D:wrist;E.hip; F.knee; G.ankle. (see Fig 3)

This algorithm is tested by a 400m dash image sequence(15 frames) with real world background and illumination. The image resolution is 512 by 512. The feature points on the first three frames are carefully located by hands. Fig 1 gives experiment results. As it has shown trajectories of A,B,C,E,F,G are all quite satisfactory. But trajectory D is notreliable because the wrist shape is so blur and its grey-scale is so close to the backgroud even in the initial frames. We also import one human interfare on trajectory C in frame 8, because the elbow sharp changes greatly from frame 7 to frame 8.

As a whole, this algorithm is quite robust even in very complicated real world circumtances. With the combination of motion smoothness and grey-scale similarity assumption, the algorithm chooses feature points, which are imposed by humaninterests and not concern object properties at all, in concective frames and forms a trajectory by these correspondingpoints.

Often when sharp deforms in the successive frame, some unexpected results tend to occure. Research which is ongoing will be focused on this problem with some flexible matching methods.

[1]N.Suwa, N.Sugie and K.Fujimura, "A preliminary note on pattern recognition of human emotional expression", Proc. of the 4thIJCPR, 1978, pp408—410.

[2]K.Arita, "Image sequence analysis of real world human motion", P.R., vol17, No.1, 1984, pp73—84.

[3]Lu Qin and Zhou Xin, "Window tracking technique on analysingreal world human body motion", Proc. of workshop on Machine Vision Application, 1990, Japan, pp236—241.

[4]Maylor K.Leung and Yee-Hong Yang,"Human body motion segmentation in a complex scene", Pattern Recognition, Vol.22, No.1, pp55—64, 1987.

[5]I.K.Sethi,V.Salri and S.Vemuri, "Feature point matching in image sequence", Patter Recognition Letter 7, No.2, pp113—124, 1988.

[6]B.Widrow, "The rubber mask technique—I.Pattern measurement and analysis", Pattern Recognition, Vol 5, pp175—198, 1973.

[7]I.K.Sethi and R.Jain, "Finding trajectories of feature points in a monocular image sequence", IEEE Tras. PAMI-9, No.1, 1987, pp56—73.

[8]C.L.Cheng and J.K.Aggarwal, "A two-stage hybrid approach to the correspondence problem via forward -search and backward-correction", in Proc. of ICPR-A, 1990, pp173—178.

[9]Martin D.Levin, "Computer determination of depth-maps", Computer Graphics and Image Processing, Vol 2,1973,pp131—150.

[10]Ramakant Nevatia, "Depth measurement by motion stereo", Computer Graphics and Image Processing, Vol 5,1976,pp203—214.

[11]Saburo Tsuji, Michiharu Osada and Masahiko Yachida, "Tracking and segmentation of moving objects in dynamic line images", IEEE Trans.PAMI-2, No.2, 1980, pp516—521.

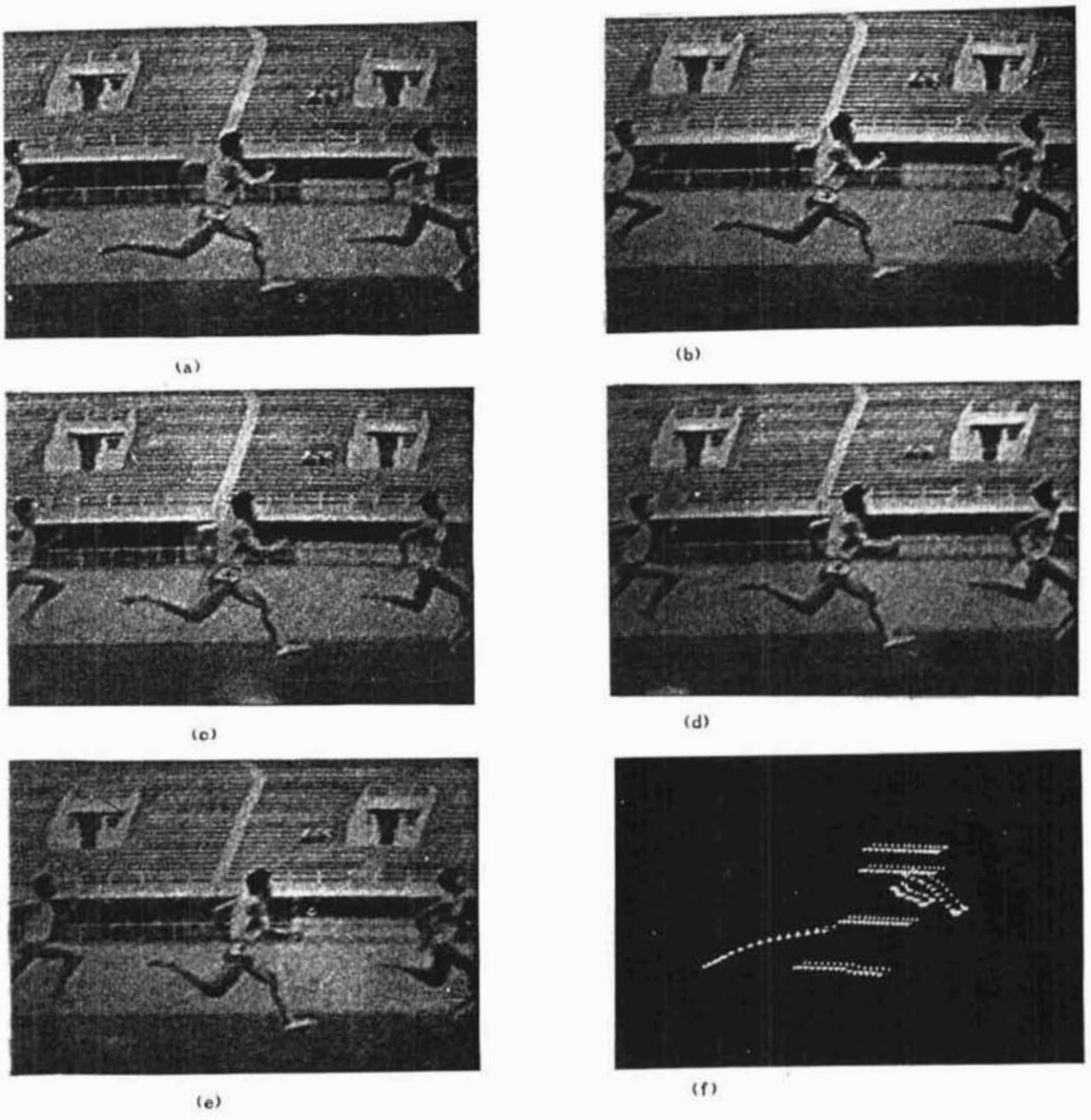

Fig 1. Experiment result: tracked feature points in sequential images and their trajectories. (a)Frame 1; (b)Frame 4; (c)Frame 7; (d)Frame 11; (e)Frame 15; (f)Trajectories of seven feature points.

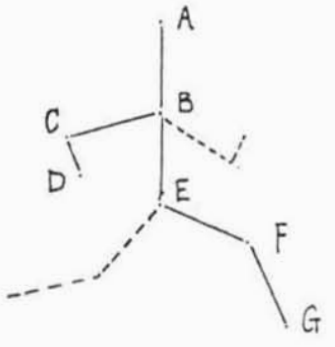

Fig 3. Skeleton of A Running Person: A.Head; B.Should; C.Elbow; D.Wrist; E.Hip; F.Knee; G.Ankle.
…… unvisible side.
—— visible side.

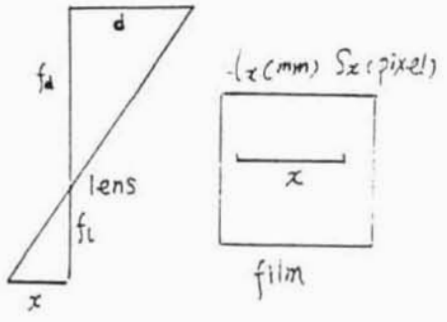

Fig 2. Estimation of disparity X:

$x = d*f_1/f_d(mm) = s_x*d*f_1/(l_x*f_d)(pixel)$