MVA'90 IAPR Workshop on Machine Vision Applications Nov. 28-30,1990, Tokyo

IMAGE SEQUENCE ANALYSIS OF REAL WORLD HUMAN BODY MOTION

Xin ZHou,Qing LU,Zhe GU
Department of Computer Science,Fudan University

220 Handan Road
Shanghai,200433,P.R.China

ABSTRACT

In this paper attempts have been focused on using window-tracking technique to analyse human body movement.According to the behavior of human vision system,an idea of rough and precise search-space is proposed in order to reduce the search-space.Some properties of windows which fit tracking the joint movements of the real world human body are illustrated.Further discussion on window-tracking technique is presented at last.

INTRODUCTION

To analyse human body motion ,traditional method is to indicate the locations of joints in each image frame manually,then calculation of speeds,acceleration and trajectories is done by computer.It needs great amount of work but the result is not very precise.Another method is to put some marks on the places to be investigated,then analyse their movements and change,such as recognition of human emotional expression done by N.Suwa and others. Besides, Koichiro Akita adopts a method of segment and representation under the conduction of human body model. But his model is too simple, and the interrelationship between images in the sequence is not utilized, so this method can't be used widely.

In this paper we adopt a window-tracking technique to analyse human body movement.It can be used to analyse human body motion in real world.

WINDOW TRACKING TECHNIQUE

1. General window-tracking technique: General window_tracking technique is adopted in tracking rigid body motion. As shown in Fig.1,when rigid body moves from location 1 to location 2,the window also tracks from location a to location b. We express the technique with formula as follow:

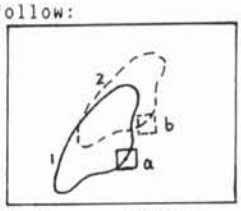

Fig.1 Window-Tracking rigid body

Fig.2 Human body model

Define the second frame as f(x,y),
Define the first frame as w(x,y),
then tracking evaluation function is:

$$R(m,n)=\sum_x \sum_y [f(x,y)-w(x-m,y-n)]^2$$

Obviously,R(m,n) is minimum if the window content of first frame matches that of second frame.That is to find (m_o,n_o) which satisfied:

$$R(m_o,n_o)=\min_{(m,n)\in s} R(m,n)$$

where s is search-space.

2. Human Body Model:The human body is represented as the combinatorial body of rigid bodies approximately.In this case,a joint is a unique point connecting two rigid bodies.So the movement of human body can be described if the joints can be tracked.Because postures of human body is mainly described by 13 joints,and many human body movements, such as walking and running, are symmetry,so we only need to track 7 joints in our experiment.They are center of head and joints on shoulder ,elbow ,wrist, hip, knee and ankle,as shown in Fig.2.

3. Actual Analysis Technique:

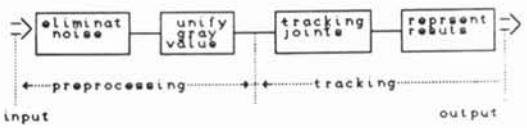

Fig.3

3.1 Overview: Fig.3 is the block diagram

of the analysis system.The whole process is composed of two stages: preprocessing and tracking.In first stage,noise is eliminated from each frame by wave filter,and image sequence gray value is unified so that every image frame has the same average gray value.In second stage,the window's location in present image frame is sought according to the content of the window in previous image frame,and trajectory is shown by tracking results.

TO analyse the image sequence, first the joints of the first image frame are pointed out manually by man-machine interaction,then square windows are opened at each location of joints in first frame by computer ,so as to track the joints in next frame .The content of the windows changes relevantly during the procedure.

3.2 Preprocess

1)Smoothing and Eliminate Noise:We adopt fast middle-value filter to eliminate noise of inputed images.The structure of the filter is shown as Fig.4

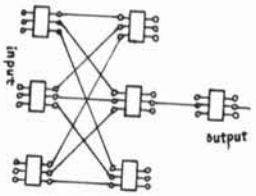

Fig.4

2)Gray Value Uniformation: As the images' brightness of each frame can't be absolutely the same,it's possible that the object has different gray value in different frame.So it's neccessary to unify the gray value of each image frame.

Suppose there are M frames in a image sequence,and each image is made up of n*m pixels(in our experiment,n=512,m=512). $f_k(i,j)$ is the gray value of the kth image,where k=0,1,...M-1; i=1,2...m; j=1,2...n. Then the average gray value of each frame is:

$$avg_k = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} f_k(i,j)$$

The second and the next images are processed as follow:

step 1: $f_k(i,j) = f_k(i,j) + avg_o - avg_k$

where x is the take-down integer

step 2:

1. Generating random number ω in (0,1)

2. $$f_k(i,j) = \begin{cases} f_k(i,j)+1 & \text{if } \omega \le [avg_o - avg_k] \\ f_k(i,j) & \text{otherwise} \end{cases}$$

where $[x] = x - x$;k=1,2...M-1.

3.3 Tracking

Window-tracking is the technique which search location of the window in succeeding image by matching the content of the window in previous image. That is to find coordinate point (m_o, n_o) in succeeding image frame which satisfied formula (2). The window size and tracking tactics should be considered.

1) Window Size:As the background of the image is rather complex and may be different in different frame,what's more,the surface of human body is smooth,and the gray value of the pixels is close together,so if the window size is too big,and much background is enclosed ,tracking can't be precise.But if the window size is so small that the window is in the area of human body,boundaries and outlines which cause big difference of gray value will lose.We get two rules by experiment.The window should include more outlines and boundaries so that the gray histogram of window content has much difference .At the mean time,the window should enclosed less background.In fact the rules are contradictory,so we have to consider trade-off when we choose the window size.

We set different window on different joint considering its distinct features.In chart.1,Lx,Ly each stands for the window size on x-axis and y-axis.

Chart.1 The Window Size for Defferent Joint

head		shouder		elbow	
lx	ly	lx	ly	lx	ly
20	20	28	20	28	28

wrist		hip		knee		ankle	
lx	ly	lx	ly	lx	ly	lx	ly
20	28	20	28	28	28	20	20

2)Tracking Tactics: As the background of human body motion images is real world,pure tracking methods such as gradient methods can't be used in this case.And because complex background may cause several minimum values in search space,global search is neccessary.Based on it ,the least is found as the result of tracking. There are two constraints to reduce the search space.

Distance Constraint:The displacement of a point on the motive 4 object between

succeeding images is definite,because the interval τ between two frame is very short. Velocity Constraint:Moving object's velocity won't change much in τ as a result of interia.So the range of a point's relevant location in succeeding image can be estimated according to the previous velocity of the object.

Suppose finished sequent frames are f_{k-1}, f_k ,present frame is f_{k+1}; A point on moving object has coordinate (x_{-1},y_{-1}),(x_0,y_0) and (x,y) on f_{k-1},f_k and f_{k+1} separately;Distance constraint is t_d ,velocity constraint is t_v .Then search space s_h is

$$S_h=S_{h1}\cap S_{h2}$$

where $S_{h1}=\{(x,y)\mid (x-x_0)^2+(y-y_0)^2\le T_d^2\}$

$$S_{h2}=\{(x,y)\mid (x-2x_0+x_{-1})^2+(y-2y_0+y_{-1})^2\le T_v^2\}$$

When people want to find an object,usually they search its location roughly in a big space,then take a further step to decide the precise location.We accept this idea in searching technique,adopt a two-step method to reduce the search space.

Step 1:Rough Search and Match

Rough location can be obtained by scanning the image on interlacing column and row.So rough search space can be defined as follow:

$$S_a=S_{a1}\cap S_{a2}$$

where $S_{a1}=\{(2x,2y)\mid (2x-x_0)^2+(2y-y_0)^2\le T_d^2\}$

$$S_{a2}=\{(2x,2y)\mid (2x-2x_0+x_{-1})^2+(2y-2y_0+y_{-1})^2\le T_v^2\}$$

So rough matching is:

$$R(m,n)=\min_{(k,l)\in S_a} R(k,l)$$

Step 2:Precise Search and Match:

Precise search is correcting the result of rough search.Suppose the approximate location given by rough saerch is (m,n),then precise search space S_p is made up of (m,n) and its nearest pixels.

$$S_p=\{(x,y)\mid (|x-m|\le 1)\wedge(|y-n|\le 1)\}$$

So,precise search and match is

$$R(m_0,n_0)=\min_{(m,n)\in S_p} R(m,n)$$

In the section follows,we compare the time cost by two-step search and match with that cost by global search and match mathod.

suppose one operation of search and match spends time t,then the ratio of two methods' time consumtion is :

$$\sigma=\frac{\sum_{(x,y)\in S_a} t+9t-t}{\sum_{(x,y)\in S_h} t}=\frac{\sum_{(x,y)\in S_a} 1+8}{\sum_{(x,y)\in S_h} 1}$$

Obviously,

$$\lim_{\|s_h\|\to\infty}\sigma=\frac{1}{4}$$

Generally,s_h is rather small,then $\sigma\sim\frac{1}{2}$,that is, search time can be reduced to nearly half of global search time.

CONCLUSION

Here we review some problems that require futher discussion.

The technique can be improved in two respects.First,the major content of the window is a part of the object being tracked,and the surface of the object(human body parts) is rather smooth.So generally the gray histogram will appear as Fig.5,where there are 1--n frequency peaks which are obviously different from the others. Their correspondent gray value section is the range of the gray value of the object being tracked.So,we can eliminate background in the window by mapping other gray value to a single gray value.Second,the window tracking technique will fit for tracking the movement which changes in the direction of depth by adding scale factor.

But still there are some problems.The first one is that images should not have much noise because tracking and matching is done in the light of gray value,otherwise the results won't be precise. Another problem is that this technique can't track the object invisible in some frames of the sequence.

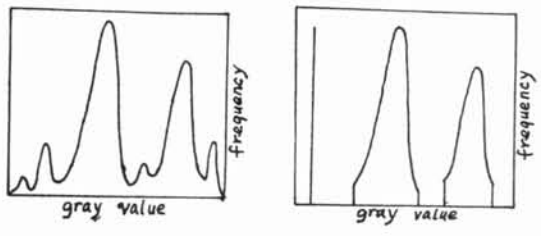

Fig.5

ACKNOWLEDGMENT

The support of the National Foundation for Natural Sience is gratefully acknowledged.

REFERENCES

[1] N.Suwa,N.Sugie and K.Fujimura,"A Preliminary Nate on Pattern Recognition of Human Emotional Expression",Proc.of the 4th IJCPR, 1978, pp408-410.

[2] K.Akita,"Image Sequence Analysis of Real World Woman Motion", P.R, Vol.17, No.1, 1984, pp73-83.

[3] T.S.Huang and R.Y.Tsai,"Image Sequence Analysis: Motion Estimation",Image Sequence Analysis (ed. by T.S.Huang),Spring-Verlag Berlin, 1981, pp1-36.

[4] S.Ullman,"Recent Computational Studies in the Interpretation of Structure from Motion",Human and Machine Vision,MIT Press, Cambridge, Masschusetts, 1981, pp459-479.

[5]M.K.Leung and Y.H.Yang "Human Body Motion Segmentation in a Complex Scene," P.R., Vol.20, No.1, 1987, pp55-64.

[6]M.K.Leung and Y.H.Yang,"A Region Based Approach for Human Body Motion Analysis," P.R., Vol.20, No.3, 1987,pp321-339. 7