MOTION ESTIMATION AND DETECTION OF COMPLEX OBJECT BY ANALYZING RESAMPLED MOVEMENTS OF PARTS

Punpiti Piamsa-nga[1] *and Noboru Babaguchi*[2]

[1] Department of Computer Engineering, Kasetsart University,
Jatujak, Bangkok 10900 THAILAND (pp@ku.ac.th)

[2] Department of Communication Engineering, Osaka University,
2-1 Yamadaoka, Suita, Osaka, 565-0871 JAPAN (babaguchi@comm.eng.osaka-u.ac.jp)

ABSTRACT

A moving object that has many complex moving parts is very hard to detect and its motion is not easy to estimate. In this paper, we present a new technique for motion estimation and detection of moving complex objects by analyzing the resampled motions of the parts of objects. The Kalman filter is used to track all resampled movements and the tracked routes are classified into groups that share the same fundamental movements. Our simulation show that recall of motion estimation and detection is approximately 0.8, while the computation drops exponentially.

1. INTRODUCTION

In the field of image/video processing and analysis, object detection is a very important tool since it is regularly used in the first step in complex video processing. Many techniques have been proposed for a decade with various aspects; however, we found that the most popular techniques among those are the blob analysis of unknown objects and the articulated motion object analysis.

The blob analysis is to detect object motion by recognizing a moving object as a blob; therefore, all movements of its parts inside the blob are intentionally ignored. For example, in a scene of a human walking, blob analysis of human walking usually ignore the swing movements of hands and legs. In other words, we can say that the analysis is done at a particular resolution of the movement. The blob analysis is very useful in various applications, such as security surveillance, pedestrian and traffic control, etc [2][3][4][5].

On the other hand, analysis of articulated objects is an opposite approach to the blob analysis since it is used for recognizing objects whose all possible motions are completely known. In this case, the specifically interesting movements are used in the analysis and uninteresting movements are ignored. As a result, it is then usually used in specific applications in close environments, such as industrial robot arm detection [6].

To determine the motions of the objects, we found that the Bayesian techniques, especially the Kalman filter, have been frequently exploited [2][3][4][5][6][7]. Whether the possible motions can be described or not, the results of Kalman filter for motion detection and estimation are very useful and successful with a condition that possible movements of objects are assumed to comply with linear systems. In this case, all random movements and occurrences are ignored.

In the general case, in order to detect an unknown moving object that attaches lots of complex moving parts, we do not know the correct resolution of the motion before we see it. Also, both blob analysis and articulated object analysis are appropriate for some but all applications. In the real-world of movements, there are lots of parts where their movements are independent but they are actually connected visibly or invisibly and they move together. For example, while a man walking, movements of arms and legs may seem independently but they are still connected or while many soldiers marching, they all have a connection as a troop. A human can easily find that hands are attached to the body and soldiers are in the same troop by considering at a different resolution of the movement. In other words, we can use the resampled motions in order to detect a complex object which has multiple moving parts.

In this paper, a simple technique for detecting moving complex objects in video data using resampled movement estimation of their parts is proposed. Our approach is presented in Section 2. The mathematical model and the algorithm are described in Section 3. The experimental results are illustrated and explained in Section 4. The conclusion and our future work are given in Section 5.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

2. THE APPROACH

We propose motion estimation and object detection technique for an articulated object, which has many complex moving parts attached, in video stream data. Our approach is adapted from a simple, natural behavior of human. Naturally, a human considers object movement globally in video frames within a reasonably long period of time in order to identify the direction of an object. On the other hand, a human considers a specific area in a shorter period in order to investigate local movement. For example, in a scene of human walking, if we see in a shorter time and more specific area, we can recognize a swing pattern of arms and legs and steady movement of body; then again, if we look in a wider area within a longer period, we are able to recognize his/her moving destination, while the motion patterns of arms and legs look like are not necessary to be in consideration at all. Since the movement of arms and legs has the same fundamental movement as that of the body, we can recognize that arms, legs and body are composed of as a moving complex object [1].

Technically speaking, we resample motions in time domain at a lower sampling frequency in order to find a global movement of objects and then we use the global movement to classify where the objects are.

3. THE MODEL

3.1. The mathematical model

We assume that the motions of objects must follow the linear time-invariance system properties; therefore, moving information in a particular frame can be derived from the previous ones. The model of the moving route can be defined by:

$$\begin{aligned} \dot{x}[n] &= Ax[n] + Bu[n] + w[n] \\ y[n] &= Cx[n] + v[n] \end{aligned} \tag{1}$$

n is an order in a discrete-time sequence. $x[n]$ is the state variable – generally $x[n]$ is a vector representing positions in coordinates, speeds and accelerations along all axes. $u[n]$ is an input vector; A , B , and C are coefficient matrices, which describe the characteristics of the movement; $v[n]$ and $w[n]$ are process and measurement white noises with constant covariance, respectively.

Suppose a moving complex object travels on a finite moving route $f[n]$ between time interval $[N_1, N_2]$ and it composes of R moving parts. Then, there are different routes $x_r[n]$, where $r \in [1, R]$. Since every part is a part of the same moving complex object, all routes $x_r[n]$ must share the fundamental route $f[n]$. Let $p_r[n]$ be the additive motion of moving part to the fundamental route. Then,

$$p_r[n] = x_r[n] - f[n] \tag{2}$$

Since each moving part must have the same destination, if the interest time interval $(N_2 - N_1)$ is appropriately long, we can assume that, for $r \in [1, R]$ and $n \in [N_1, N_2]$,

$$average(x_r[n]) = average(f[n]) \tag{3}$$

$$\forall n, average(p_r[n]) = 0 \tag{4}$$

Within the interval $[N_1, N_2]$, if the sampling frequency onto the route $x_r[n]$ is not lower than the frequency of the fundamental movement, the Equation (4) is a sufficient condition to identify that all $x_r[n]$ share the fundamental motion in the $[N_1, N_2]$ interval. The moving complex object detection can be done in these following four steps. First, since $p_r[n]$ and $f[n]$ are unknown, to find a condition of Equation (3) can be done by finding all $average(x_r[n])$ in an appropriately-selected, long time interval. Second, the Kalman filter is exploited to estimate all the routes. Third, each estimated route is compared with all the rest to generate a distance matrix. Finally, a clustering algorithm, such as k-mean (also used in this paper,) is exploited to classify which routes share the same fundamental motion.

Practically, this computing process takes long time; however, the condition in Equation (3) is true only if the selected time is long enough. Consequently, if we need a more correct result, then it need to compute longer

To meet the condition in Equation (3) with reducing the cost of high computation can be done by resampling the routes at a lower sampling frequency. Note that, according to Equation (3), the averages of original sequence and those of the resampled sequence must be "theoretically" the same if we consider it in appropriately long observation time. Therefore, the resampling approach can not only extend the observation period which improve the correctness of motion estimation but also reduce the computation cost in the route classification.

3.2. The algorithm analysis

In this Section, we analyze the computation of our model by the analysis of the algorithms, described in Figure 1. We can notice that the computation cost depends on a length of route $(l = N_2 - N_1 + 1)$, the number of routes (R) and the number of decimation (d) . Therefore, time complexities of the Kalman filter for one route and clustering process are $O(l/d)$ and $O(R^2/d^2)$, respectively. Group comparison complexity, which depends on the number of members in the cluster, is small; therefore, we omit it. Note that if the sampling rate decreases (or the d increases), the computation cost decreases. In other words, the total computation depends on the number of decimation, which is arbitrary between one and half of total original sampling.

```
$d$ =2;
PreviousGroup:=NULL;
while $d$ < half a length of route($l/2$)
    for each route $r$
       Resampled[$r$]:=Decimate(Original,$d$);
       Estimated[$r$]:=Kalman(Resampled[$r$]);
    end for
    Group[$\cdot$]:=Clustering($\forall r$,Estimated[$r$]);
    Distance:=Compare(Group[$\cdot$],PreviousGroup);
    PreviousGroup:=Group;
    $d$ :=$d$ +1;
    if (Distance<THRESHOLD)
      return;
    end if
end while
```

Figure 1: The proposed algorithm

4. EXPERIMENTS

4.1. The experiment environments

The experiments were performed under a simulated environment. We assumed that the observation area is fixed (fixed camera environment) in a two-dimensional grid space through the observation period. We considered only the moving complex objects that do not break into pieces and scatter into drastically different directions within the observation period. Since the length of observation period is arbitrary, we chose ten seconds long. The original sampling rate is selected at 25 frames per second. Therefore, the total number of sampling of a route is 250 and time interval between each sampling is 1/25 second. The motion estimation will not be possible if the route is out of the observation area longer than a sampling period.

All movements are on a 2-D plane and they are assumed to be complied with the Newton laws in order to assure that the movement can be possible in real situation. First, the fundamental movement is generated, where averages and variances of speeds and accelerations, and the constant-covariance measurement and process noises on both X- and Y-directions are the parameters. Second, since all moving parts have the same fundamental movement, we put all parts on to the fundamental route and inject an additive movement to each part. At each part, an observed point and a center-of-gravity point (CG) are assigned. All parts are set to spin around its CG at a given angular velocity, where CG is always on the fundamental route. Finally, the motions of observed point on each moving part are used as routes in the experiments. An example of route is depicted in Figure 2.

The MATLAB 6 R13 on a Linux PC is exploited for all simulations. Recall and computation time are used as measurement metrics of our approach.

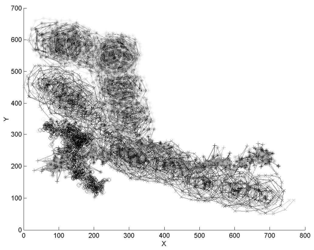

Figure 2: An example: original routes

4.2. Motion estimation and detection

We changed the sampling rate by decimation from 1 to 32 points, iteratively. After that, we used them to estimate the motion by the Kalman filter. The estimated motion was then transformed to a sequence of moving direction (in radian) and a sequence of directional speed (in distance unit/second.) Figure 3 shows the directions of each estimated route at selected re-sampling frequency from 1/4 to 1/32 of original route in Figure 2. Each legend in the graph is a route of an object. Note that, behavior of route estimation in Figure 3 is extremely unreadable when the low number of decimation (4 and 8,) is used; in contrast, at the large number of decimation (16 and 32,) it is easy to classify similar tracks into corresponding groups.

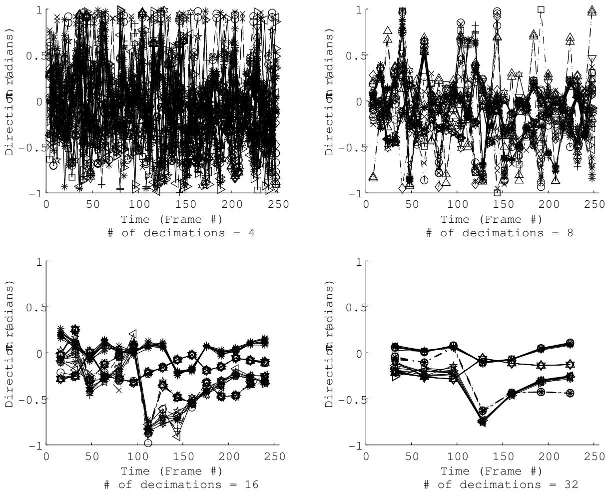

Figure 3: Examples of motion estimation from resampled routes at selected number of decimations

Both direction and speed sequences are used for classifying the routes that have similar or slightly different movements into a group by the k-mean clustering algorithm. The classified routes are shown in Figure 4. At higher number of decimations, it is easier to determine the fundamental movement of moving complex objects.

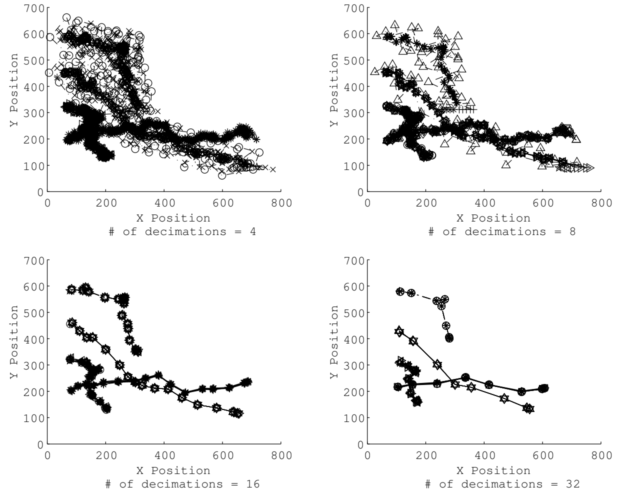

Figure 4: Examples of classified routes of moving complex objects in Figure 3 at selected number of decimations

4.3. Measuring effectiveness of the algorithm

Recall and comparison times are depicted in Figure 5. We generated 1000 fundamental-motion routes and each moving complex object composed of 10 moving parts. The generated route and the composition of articulated route were kept as the "answer." The generated routes were resampled by decimation from one point to 32 points and then sampled routes were classified by k-mean clustering algorithms. Next, the results from the clustering process were compared with the "answer." If a route is clustered correctly, it is called as "detection". If not, it is called as "miss." Then, the recall is defined by:

$$recall = \frac{\text{number of "detection"}}{\text{number of "detection"} + \text{number of "miss"}} \quad (5)$$

Figure 5 shows that the recall is approximately 0.8 and there is no significant change of recall when the number of decimations changes; on the other hand, the computation (number of comparisons) decreases exponentially.

5. CONCLUSION

We have proposed a new technique for motion estimation for moving complex objects that their parts have common movement by changing the sampling frequency in order to find a possible fundamental movement. We have presented the idea in both mathematical model and algorithms. The experimental results show the effectiveness of our techniques both graphically and quantitatively as well. There is no significant difference of recall of motion estimation at different sampling frequency but the computation time decrease exponentially.

Our future work is to extend the observation period to general period and find a more suitable model to represent a fundamental movement. We believe that it would be very useful for various applications if a multiresolution approach also could be applied into our model.

6. REFERENCES

[1] Y.-M. Chou and H.-M. Hang, "A new motion estimation method using frequency components," *J. of Visual Communication and Image Representation* 8:1 83-96, 1997.

[2] E. Stringa and C. S. Regazzoni, "Real-time video-shot detection for scene surveillance applications," *IEEE Trans. on Image Processing,* 9:1, 69-79, 2000.

[3] D.-C. Chang and W.-R. Wu, "Maneuvering target tracking with high-order correlated noise – a multirate Kalman filter approach," *Wireless Personal Communications,* 17, 102-123, 2001.

[4] C. Sacchi, G. Gera, L. Marcenaro, and C. S. Regazzoni, "Advanced image-processing tools for counting people in tourist site-monitoring applications," *Signal Processing,* 81, 1017-1040, 2001.

[5] O. Masoud and N.P. Papanikolopoulos, "A novel method for tracking and counting pedestrians in real-time using a single camera," *IEEE Trans. on Vehicular Technology,* 50:5 1267-1278, 2001.

[6] K. Nickels, "Model-based tracking of complex articulated objects," *IEEE Trans. on Robotics and Automation,*17:1, 28-36, 2001

[7] P. Sykacek and S. Roberststs, "Bayesian time series classification," *Advances in Neural Information Processing Systems*, 14, 2001

[8] J.F. Lynch, "Analysis and application of adaptive sampling," *J. of Computer and System Sciences,* 66, 2-19, 2003.

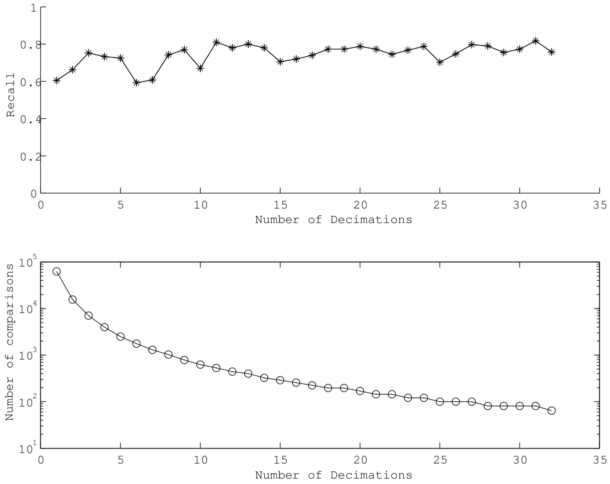

Figure 5: Recalls and number of comparisons