The Use of GTFR with Cone Shaped Kernel for Motion Estimation

Vinay G. Vaidya
Robert M. Haralick

Intelligent Systems Lab.
Dept. of Electrical Engineering
University of Washington
Seattle, WA 98124, U.S.A.

Abstract

This paper presents a new method for estimating 2D motion parameters from a noisy 2D image sequence. Our performance characterization is based on 20,000 simulated noisy images. The results show that even with a signal to noise ratio of 5 db, one can detect rectangular objects of at least 4 pixels in length and width, to within 2.5 pixel location accuracy. The misdetection rate is near zero and the average false detection rate is 0.06 false objects per frame of 64 X 64 size in a 5 db SNR environment. With -1 db SNR we can detect objects within 5 pixel accuracy. The misdetection rate for this case is near zero and the average false detection rate is 1.6 false objects per frame of 64 X 64 size. The method is also applied to an image sequence obtained from a 747 takeoff scene. Using the new method, the 747 takeoff speed was predicted to be 142 kts, which is well within the typical range of 140 to 150 kts. The results presented here clearly show successful application of the method to real dynamic scenes.

1 Introduction

Motion estimation is one of the main research areas in computer vision. In order to solve the motion estimation problem, a number of methods have been proposed. Most of these methods assume CAD model of the objects in motion or assume point correspondence. If neither of these are available then the choice of methods is restricted. One can use the method of differentials and subtract two images to estimate motion parameters. However, the approach breaks down in presence of noisy environment. Another approach is to use the Fourier transforms. This method, however, is restricted to single object motion, and constant, noise free background (Huang [3]). The third approach is to use the optic flow method. The performance of this method also degrades with noise.

One can use the Wigner distribution to identify object locations. These objects can then be tracked to find estimates of the motion parameters. The Wigner distribution however suffers from effects of the cross terms. These cross terms in our application imply presence of objects which are pseudo objects. In this paper we present a method that virtually overcomes the effect of cross terms and helps identify object locations. The method uses the generalized time frequency representation (GTFR) with the cone shaped kernel.

2 Problem Statement

Consider a scenario where we have no knowledge of the object sizes or shapes. Our goal is to estimate motion parameters from a sequence of images taken from a complex dynamic scene. In this scenario, there are multiple objects moving in different directions, with different velocities. The image sequence is corrupted with uncorrelated noise. The object motion takes place in a 2D plane which is parallel to the plane of the image. These images of the objects are orthogonal projections. It is further assumed that in a given sequence, objects do not occlude one another.

The images are assumed to be grey scale images with grey scale values greater than zero. It is further assumed that the objects of interest are brighter than
the background. In cases where this condition can not be met, the images can be inverted in grey scale.

3 Generalized Space Frequency Distribution Functions

The generalized class of time frequency representations was developed by L. Cohen [1]. In discrete domain [2] the GTFR for a function $f(t)$ can be written as

$$C_f(t, u; \phi) = 2 \sum_{t' = -M}^{t + \frac{M}{2}} \sum_{k = -M}^{k} \phi(t - t', k)$$

$$f(t + k)f(t - k)e^{-j\frac{2\pi}{M}uk}$$  \hspace{1cm} (1)

where $M$ is even, and $\phi$ is an arbitrary kernel function that determines a particular class of representation.

If we change time variable $t$ to space variable $x$, and $t'$ to $p_x$ then the equation (1) can be written as

$$C_f(p_x, u; \phi) = 2 \sum_{x = -\frac{p_x}{2}}^{p_x} \sum_{k = -\frac{p_x}{2}}^{k} \phi(p_x - x, k)$$

$$f(x + k)f(x - k)e^{-j\frac{2\pi}{M}uk}$$

This can be further extended to a 2 dimensional function $f(x, y)$ as

$$C_f(p_x, p_y, u, v; \phi) =$$

$$2 \sum_{x = -\frac{p_x}{2}}^{p_x} \sum_{y = -\frac{p_y}{2}}^{p_y} \sum_{k = -\frac{p_x}{2}}^{k} \sum_{l = -\frac{p_x}{2}}^{l} \phi(p_x - x, p_y - y, k, l)$$

$$f(x + k, y + l)f(x - k, y - l)e^{-j\frac{2\pi}{M}(x + y)} (2)$$

This equation now can be used for image processing applications. In this case $N$ is the number of rows in an image and $M$ is the number of columns.

4 Cone Kernel Derivation

Zhao, Atlas, and Marks [4] derived the relationship for a cone shaped kernel for 1D signal and used the derived equations for speech processing. We have extended their work for 2D signals. In the 2D signal case the correlation function becomes a function of 4 variables and is given by

$$g(x, y, k, l) = f(x + k, y + l)f(x - k, y - l)e^{-j2\pi(\frac{x}{M} + \frac{y}{M})}$$

The cone shaped kernel, denoted by $ZAM(p_x, p_y, k, l)$, can be written as

$$ZAM(p_x, p_y, k, l) = \sum_{x = p_x - k}^{p_x + k} \sum_{y = p_y - l}^{p_y + l} g(x, y, k, l)$$

The above equation when computed for $k = a$ and $l = c$ becomes

$$ZAM(p_x, p_y, a, c) = \sum_{x = p_x - a}^{p_x + a} \sum_{y = p_y - c}^{p_y + c} g(x, y, a, c)$$

After substituting the equation for $ZAM$, the equation (3) becomes

$$C_f(p_x, p_y, u, v; \phi) = 2 \sum_{k = -\frac{p_x}{2}}^{p_x} \sum_{l = -\frac{p_x}{2}}^{l} \phi(p_x - x, p_y - y, k, l)$$

$$ZAM(p_x, p_y, k, l)$$  \hspace{1cm} (3)

The limits in equation would however depend upon the selected window $q(x, y)$. This window is selected in such a way that the cross term effects are nullified. Fig. 1 shows the cone kernel in a geometric form. All points within the cone are summed up.

5 Choosing the Window Size

The equation (2) for $C_f$ is computationally quite complex. The maximum value for $k = \frac{M}{2}$ and for $l$ it is $\frac{N}{2}$, where $M$ is the number of columns and $N$ is the
number of rows. The number of summations in the equation depend upon these maximum values. The values can be restricted based upon how one selects the window $q(x,y)$. This window selection is also based upon how many neighborhood pixels does one want to take into account. Larger the window more smoother the function would be.

Another important aspect to consider is the elimination of pseudo peaks or cross terms. If the window is large, the chances of two signal interacting and giving rise to the cross terms increase. If the window is smaller than the distance between the two signals then the cross terms will not arise. Therefore, restricting the window size would benefit from two different points of view. First, it would eliminate or reduce cross terms and secondly, computationally it would be less expensive.

6 Example

Consider a simulated image of two rectangular objects shown in Fig. 2. We choose the limits on $k$ and $l$ to be 3 which gives 7 pixel wide window. The frequencies $u$ and $v$ are selected to be zero. The result of using the cone shaped kernel is shown in Fig. 3. When the image is operated by the cone kernel then the output shows peaks corresponding to the object location. The main idea is to track these peaks and find out motion parameters. In Fig. 3, we see only two peaks indicating that there are two objects. Thus, there are no pseudo peaks. If a pseudo Wigner distribution is used it would result in 3 peaks out of which two are true peaks and the third one would be a pseudo peak. Fig. 4 shows the same image corrupted by Gaussian noise with -1 db SNR. Fig. 5 shows the result of using cone shaped kernel. Although the output is noisy, one can easily discriminate between the noise and the objects which is not at all possible in the original noisy image. This example shows how one can effectively eliminate cross term effects using the cone shaped kernel.

7 Performance Characterization

The performance characterization of the new method is based on 20,000 simulated images. The simulated images were 64 X 64 in size and each image had 3 rectangular objects in motion. Independent Gaussian additive noise with SNR varying from -1 db to 30 db was added to the images. Fig. 6 shows two plots. The first one is error in locating object centroid vs SNR. In this plot there are 3 curves. The
top curve shows $+\sigma$ standard deviation and the bottom curve shows $-\sigma$ from the mean error. The error is averaged over 1,000 trials. It is seen that for SNR of 5 dB and higher the error is about 2 pixels. The bottom plot shows the number of centroids vs SNR. Since we had three objects in the simulated images ideally we should get 3 centroids. The average number of centroids is slightly higher than 3. In addition there are two other curves indicating $+\sigma$ and $-\sigma$ deviation from the mean centroids. The method was also applied to an image sequence obtained from a real dynamic scene. The speed of the 747 was predicted to be 142 kts.

8 Conclusion

In this paper we have shown a successful application of the GTFR with cone kernel. The results presented clearly show that this new method can detect object location even under extreme noise conditions. The results show that with a signal to noise ratio of 5 dB, one can detect rectangular objects of at least 4 pixel in length and width, to within 2.5 pixel location accuracy. The misdetection rate is near zero and the average false detection rate is 0.06 false objects per frame of 64 X 64 size. With -1 dB SNR we can detect objects within 5 pixel accuracy. The misdetection rate for this case is near zero and the average false detection rate is 1.6 false objects per frame of 64 X 64 size.

Using the new method, the 747 takeoff speed was predicted to be 142 kts. which is well within the typical range of 140 to 150 kts. The results presented here clearly show successful application of the method to real world situations.

References


