DENSE MOTION FIELD ESTIMATION BY 3-D GABOR REPRESENTATION

Mu Feng, Todd R. Reed

Department of Electrical Engineering, University of Hawaii at Manoa
muf@hawaii.edu, trreed@hawaii.edu

ABSTRACT

This paper presents two advancements in motion estimation based on the 3-D Gabor representation. First, we introduce a new searching strategy and solve the estimation bias problem encountered in the current approach to piecewise translational motion estimation; second, in order to achieve better spatiotemporal resolution, we constructed a new algorithm that can provide dense motion field estimation.

1. INTRODUCTION

The frequency domain approach to motion estimation has been of interest for a number of years. Compared with spatiotemporal methods, the frequency domain approach has many advantages. These include higher noise immunity, natural utilization of multi-frame information for complex motion models, the isolation of different motion models, and high correlation with human visual perception. A number of motion models have been found to have characteristic signatures in the Fourier domain. [4][5] However, due to the global nature of the Fourier transform, frequency domain methods cannot distinguish motions at different spatiotemporal locations. This limits the practical application of these methods.

Spatiotemporal/ spatiotemporal-frequency methods were introduced to exploit the advantages of frequency domain methods while keeping spatiotemporal locality. The 3D Gabor representation provides a convenient framework for motion analysis. The Gabor basis functions maintain optimal spatiotemporal and frequency resolution within the constraints of uncertainty. Some early results using this representation for motion analysis can be found in [1], [2] and [3].

As a novel analysis framework for motion estimation, there are many topics for further study. Principally, the frequency resolution of 3-D Gabor representation determines how precisely we can estimate the parameters of a motion model. The spatiotemporal resolution determines our ability to distinguish objects undergoing different motions. Since there is a trade off between the two abilities due to the uncertainty bound, to achieve high performance in both respects is a challenge. In the following sections, we will briefly introduce the 3-D Gabor analysis framework. Then, the existing translational motion estimation algorithms in the 3-D Gabor domain are reviewed and analyzed. We will introduce two enhancements to the algorithms: one is an improved searching strategy and the other is to solve the bias problem in velocity estimation. Finally, we develop a new algorithm to provide improved spatiotemporal resolution.

2. 3-D GABOR REPRESENTAION

As an extension of the Gabor representation of 1D signals, the 3D Gabor representation expresses an image sequence as a linear combination of a set of Gabor basis functions, localized to different spatiotemporal/spatiotemporal-frequency positions. Denote by x_0, y_0 and t_0 the spatial and temporal intervals between adjacent Gabor basis functions, and u_0, v_0 and w_0 the spatial and temporal frequency intervals respectively. The 3D Gabor basis function centered at spatiotemporal coordinates (mx_0, ny_0, kt_0) and frequency coordinates (su_0, lv_0, qw_0) can be identified by the integer index set (m,n,k,s,l,q) and expressed as:

$$gabasis_{\substack{m,n,k,\\ s,l,q,}}(x,y,t) = \hat{g}(x-mx_0, y-ny_0, t-kt_0)e^{j(su_0(x-mx_0)+lv_0(y-ny_0)+qw_0(t-kt_0))}$$

(2.1) where $$\hat{g}(x,y,t) = \frac{1}{(2\pi)^{3/2}\sigma_x\sigma_y\sigma_t} e^{-[(x/\sigma_x)^2+(y/\sigma_y)^2+(t/\sigma_t)^2\ ]/2}$$

(2.2)

A 3-D Gabor basis function is a 3D Gaussian envelope $\hat{g}(x,y,t)$ shifted to a specified spatiotemporal location and modulated by a 3D complex exponential. σ_x, σ_y and σ_t are the scales of the Gaussian envelope along the spatial and temporal axes. The basis functions are well

0-7803-8554-3/04/$20.00 ©2004 IEEE.

known for their joint spatiotemporal-frequency localization. Their 2D forms have been widely used in image processing.

Let $f(x,y,t)$ be the gray value of the pixel at location (x,y,t) in a digital image sequences, with spatial dimensions $M \times N$ and temporal dimension K. The corresponding 3D Gabor reconstruction formula is:

$$f(x,y,t) = \sum_{m,n,k,s,l,q} c_{m,n,k,s,l,q}\, gabsis_{m,n,k,s,l,q}(x,y,t) \tag{2.3}$$

where $m \times n \times k \times s \times l \times q \geq M \times N \times K$ for completeness. The Gabor transform coefficients set $c_{m,n,k,s,l,q}$ contains all the information of original signal, expressed in the 3-D Gabor framework. The Gabor coefficients associated with each basis function reflect the amount of that component at the associated spatiotemporal/spatiotemporal-frequency location. Carefully designing the sampling lattice in the joint domain, we can balance frequency and spatiotemporal resolution.

3. MOTION ESTIMATION FOR PIECEWISE TRANSLATIONAL MOTION

The piecewise uniform translation model is the simplest practical motion model. The motion trajectory is approximated as a sequence of uniform translational motions. The motion of an object is described as a set of time intervals $[t_i, t_{i+1}]$ i=0,1…K and the corresponding translation vectors V=$[V_{xi}, V_{yi}]$ representing the average velocity in the interval. If considered over an infinite time duration, the frequency domain signature of uniform translation is a plane. The slope of the plane is determined by the velocity of the translation.

In practice, we have a "cloud" of coefficients distributed in the frequency domain at each spatiotemporal location, clustered to a greater or lesser degree about a plane. The causes of this dispersion include noise and interference of neighbor motions. The algorithm we used previously was to empirically determine a threshold to remove coefficients with small magnitude and to deploy LMS estimation on the significant ones. However, selecting a threshold is highly application dependent and can affect the result greatly. Thus, we have modified the algorithm to employ weighted LMS estimation without threshold processing. It is well known that the weighted LMS solution is the best linear unbiased estimator under Gaussian white noise. Experiments show its robustness under heavy noise.

Suppose in a 3D frequency domain we have *n* frequency coefficients located at: (u_1, v_1, w_1), (u_2, v_2, w_2) … (u_n, v_n, w_n) with magnitudes $(m_1, m_2 \ldots m_n)$ respectively. They are supposed to be the solutions of a plane equation: $w_n = -r_x u_n - r_y v_n$ or $\underline{w} = -\underline{\underline{A}}\,\underline{r}$ where $\underline{\underline{A}} = [\underline{u} \quad \underline{v}]$, with reliability $f(m_n)$, where $f(m_n)$ is a positive monotonically increasing function of magnitude. In practice, we defined $f(m_n) = m_n^4$ empirically. To solve this over determined linear equation for all frequency components with respect to their reliabilities, we can construct the spatial frequency matrix $\underline{\underline{A}} = \begin{bmatrix} u_1 & \ldots & u_n \\ v_1 & \ldots & v_n \end{bmatrix}^T$, the temporal frequency matrix $w = [w_1 \quad w_2 \quad \ldots \quad w_n]^T$ and reliability matrix $\underline{\underline{M}} = diag[m_1^4 \quad m_2^4 \quad \ldots \quad m_n^4]$. The weighted LMS solution is:

$$\underline{r}_{LMS} = -(\underline{\underline{A}}^T \underline{\underline{M}}\ \underline{\underline{A}})^{-1} \underline{\underline{A}}^T \underline{\underline{M}}\ \underline{w} \tag{3.1}$$

In our previous algorithm, we used the Gabor basis functions with high spatiotemporal concentration to pursue high spatiotemporal resolution. Due to the uncertainty constraint, the frequency domain resolution decreased accordingly. With considerable uncertainty in the locations of frequency coefficients, the plane equation calculated from them may be different from the true plane to some extent, resulting in bias in determining the velocities. For example, the estimation result for velocity vectors close to horizontal and vertical directions are liable to be absolutely horizontal and vertical. By experiment, we found that velocity bias decreased quickly with increased frequency resolution. Doubling the frequency domain resolution of the Gabor basis functions yielded much better performance in velocity estimation, but with the spatiotemporal resolution halved due to the resolution trade off. In order to make up this loss, we construct dense motion field estimation by "partial reconstruction" of the image sequence. The result is a dense (pixel level) motion estimation.

4. DENSE MOTION FIELD ESTIMATION BY PARTIAL RECONSTRUCTION

In some applications such as pattern recognition in computer vision and segmentation in dynamic image processing, motion estimation is required to have high spatiotemporal resolution. In our previous algorithm, we were required to trade between the spatiotemporal resolution and frequency resolution under the uncertainty constraint. Thus, to achieve high precision in calculating velocities, frequency resolution must be rather high, sacrificing spatiotemporal resolution.

Although there is no analysis framework that can overcome the uncertainty constraint, this does not imply that information on frequency content is available only at spatiotemporal locations corresponding to basis function centers. For example, selecting a specific frequency index (s,l,q) in equation (2.3) and computing the partial sum over all (m,n,k) yields an image sequence of dimension $M \times N \times K$ corresponding to the selected frequency. Following this thought, we construct a novel algorithm that keeps the same frequency resolution as before while achieving dense motion field estimation. That is, the estimation result displays the movement status of every pixel in the image sequence.

To estimate the motion of a pixel, it is necessary to find a "local-instant" spatiotemporal-frequency representation for that pixel. To find the pixel-by-pixel frequency information, notice that the Gabor synthesis formula (2.3) can be arranged as:

$$f(x,y,t) = \sum_{s,l,q} \{\sum_{m,n,k} c_{m,n,k,s,l,q} gabsis_{\substack{m,n,k,\\ s,l,q}}(x,y,t)\} \quad (4.1)$$

where s*,* *l,* and *q* are frequency band indices, and *m*, *n* and *k* are spatiotemporal indices. We can see that the summation inside the curly braces includes all the Gabor functions with different spatiotemporal locations but the same frequency index. Actually, it is the partial reconstruction of the image sequence in a frequency band. Formula (4.1) shows that reconstruction can be done in two steps: first, partial reconstruction on each frequency band; second, sum all partial reconstruction results to form the original signal.

Each partial reconstruction result has a value at each pixel in each frame, representing the amount of signal in a specific frequency band at that spatiotemporal point. For any given pixel, collecting corresponding values from every partial reconstruction image sequence forms the "local-instant" frequency domain for that pixel.

For example: for a pixel at spatiotemporal coordinates (x_i, y_j, t_h), its frequency component at sub-band (u_s, v_l, w_h) is:

$$F_{x_i,y_j,t_h}(u_s,v_l,w_q) = \sum_{m,n,k} c_{m,n,k,s,l,q} gabsis_{\substack{m,n,k,\\ s,l,q}}(x_i,y_j,t_h) \quad (4.2)$$

Using (4.2) for all the combination of (*s, l, q*), we can have complete frequency information at the location (x_i, y_j, t_h).

Since the Gabor basis functions are localized about their spatiotemporal centers, the influences of basis functions far away from a given pixel are very small. In practice, we only include the contributions of the 26 nearest spatiotemporal neighbor coefficients in the partial reconstruction, in addition to the Gabor coefficient closest to the pixel.

The process can be briefly described as:

1) Find the 3D Gabor representation of the original image sequence.
2) For the selected pixel and frequency band, select the coefficients of the same band at the 26 nearest spatiotemporal neighbors. Together with the corresponding sub-band coefficient located closest to the pixel, compute the partial reconstruction as:

$$F_{x_i,y_j,t_h}(u_s,v_l,w_q) = \sum_{m_i-1}^{m_i+1}\sum_{n_j-1}^{n_j+1}\sum_{k_h-1}^{k_h+1} c_{m,n,k,s,l,q} gabasis_{\substack{m,n,k,\\ s,l,q}}(x_i,y_j,t_h)$$

   where (m_i, n_j, k_h) is the index of a Gabor basis function in the neighborhood of (x_i, y_j, t_h).
3) Repeat 2) for every 3D frequency band.
4) Conduct frequency domain motion estimation using a piecewise translational model and repeat 2) to 4) for every pixel of the image sequence.

5. RESULTS

Fig1 shows one of the 48 frames in an image sequence displaying three wheels undergoing counter-clockwise rotation. Each frame has 256×256 pixels and 256 possible grey levels. The noise added is white Gaussian noise with a mean of 85 and a variance of 90.

The 3D Gabor representation is computed on a $16 \times 16 \times 3$ spatiotemporal sampling lattice. Using critical sampling, we have $16 \times 16 \times 16$ frequency coefficients for each spatiotemporal location over which to calculate the weighted LMS estimation. Fig2 shows the estimation result. We can see that even with substantial noise, the methods give correct results with considerable spatiotemporal resolution.

Fig3 is the dense motion estimation result for the rectangular region in Fig2.

6. CONCLUSIONS

The 3-D Gabor representation is a prospective analysis framework for motion estimation. Its joint spatiotemporal /spatiotemporal-frequency structure takes the advantages of the frequency domain methods, with spatiotemporal locality. Doubling the frequency resolution and using a weighted LMS searching method improve the estimation performance for piecewise uniform translational motion. By calculating the “local-instant” frequency domain of a single pixel using partial reconstruction, the dense motion estimation method yields instant velocity of every pixel in an image sequence. The method provides robust motion estimation of very high spatiotemporal resolution.

7. REFERENCES

[1] T.R. Reed, “Motion analysis using the 3-D Gabor Transform” *Proceedings of the Thirtieth Annual Asiloma Conference on Signals, Systems and Computers*, Pacific Grove, California, pp. 506-509, November 3-6,1996.

[2] T. R. Reed, “On the computation of optical flow using the 3-D Gabor transform”, *Recent developments in Time-Frequency Analysis,* Kluwer Academic Publishers,1998.

[3] T. R. Reed, “A spatiotemporal/spatiotemporal-frequency interpretation of apparent motion reversal” *In Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, vol.2, pp.1140-1145, Stockholm, Sweden, July 31-Augest 6,1999.

[4] G. Cortlazzo and M. Balanza. “Frequency domain analysis of translations with piecewise cubic trajectories” *IEEE Trans. on Pattern Analysis and Machine Intelligence,*15(4):pp.411-416, 1993.

[5] G. Cortlazzo, C. M. Monti and M. Balanza. “Frequency Domain Analysis of Rotational Motion” *Multidimensional Systems and Signal Processing,* 4, pp.203-225, 1993

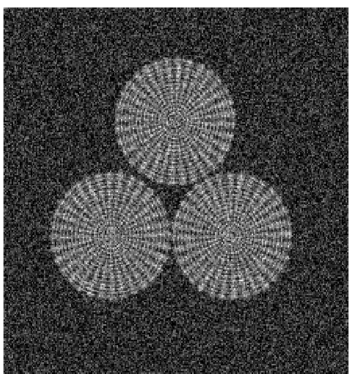

Fig1

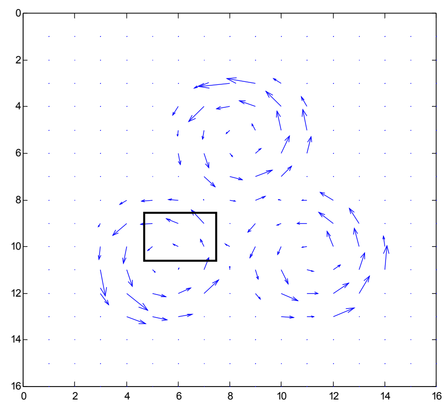

Fig2

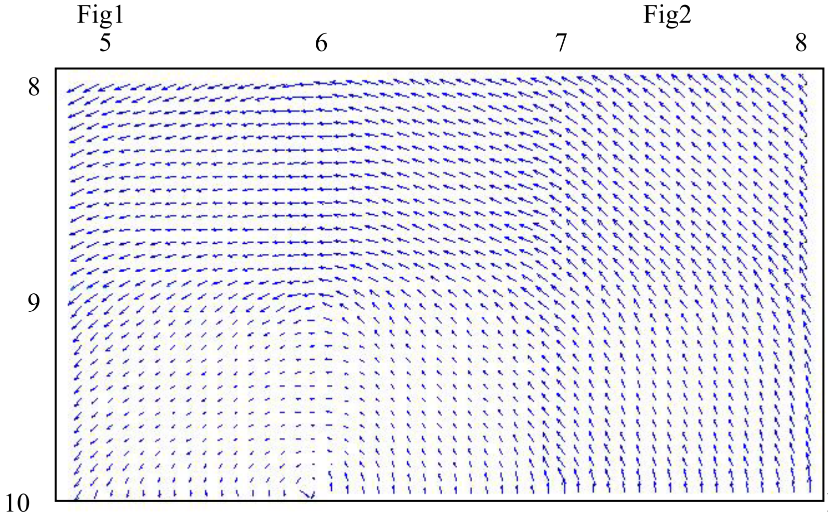

Fig3