A VELOCITY ESTIMATION METHOD FOR IMAGE SEQUENCE WITH TWO MOTION VECTORS

Shengli WU and Nozomu HAMADA

Department of System Design Engineering, Keio University
3-14-1 Hiyoshi, Kohoku-ku, Yokohama 223-8522, Japan
wu@hamada.sd.keio.ac.jp, hamada@hamada.sd.keio.ac.jp

ABSTRACT

To estimate the velocity of moving objects with arbitrary trajectory is useful in many applications. This paper aims at the velocity estimation for image sequence with two motion vectors, the problem is addressed by the transform /spatiotemporal mixed domain (MixeD) method which simplifies the 3-D signal processing problem into 1-D signal processing problem. In the appropriate spatial frequency points selected by the proposed spatial frequency selection method, the signals are not sensitive to the background and noise, therefore the corresponding spectral estimation can be addressed by FFT. The simulation results show that the proposed method is effective on the velocity estimation for arbitrary motion.

1. INTRODUCTION

In dynamic image processing, an object moving with constant velocity without changing its shape is treated as Linear Trajectory Signal (LTS), while the object with its velocity changing with time is regarded as Arbitrary Trajectory Signal (ATS). It is well known that the spectral energy of LTS is concentrated on a plane in the 3-D frequency space, and the normal vector of this plane is consistent with moving object's trajectory vector[1], therefore, the velocity of LTS can be estimated by searching this energy plane with 3-D filtering method. In the energy method, the 3-D (space-time) Gabor filters are used to sample the power spectrum of moving objects and by combining the outputs of several such filters, the slope of the energy plane (i.e., the velocity of moving objects) can be estimated[2]. But it is complicated to approach this problem in 3-D domain directly due to the multidimensionality such as the number of data.

An effective way to estimate the trajectory of LTS in the transform/spatiotemporal mixed domain (MixeD) was proposed by K.S.Knudsen et al[3]. The nonlinear trajectory estimation in the MixeD was reported in Ref.[4], it uses the chirp signal estimation technique to find the trajectory coefficients. The trajectory estimation of ATS using the adaptive trajectory estimation of LTS in the MixeD was also reported[5]. In the MixeD method, by applying the 2-D DFT to the spatial variables of 3-D signal, the problem is transformed to the problem of processing 1-D complex sinusoid signals acquired, this makes the estimation easier and faster. Usually, several spatial frequency points are selected by experience in the 2-D spatial frequency domain[3], then the spectrum of the 1-D signal is estimated by high resolution spectral estimation method. The spectral energy plane in the 3-D frequency space is determined by spatial frequency points selected and temporal frequency estimated, and moving object velocity vector can be estimated.

The track of moving objects in noisy environments based on the application of the one-dimensional Fourier transform was reported by S.A.Rajala et al[6]. In this method, all sets of pixels having the same velocity i.e. LTS is considered, and two components of the moving object velocity vector are estimated separately. It is accomplished by performing a transformation, that is by setting a weighting function, the 1-D signal can be obtained through the integration of the product of the initial signal with this weighting function, and then the 1-D DFT is used to estimate the corresponding spectrum which is in proportion to the component of velocity vector. When the weighting function is set in the form of Fourier transform, it is similar to the MixeD method.

In our study, the velocity estimation of moving objects with smoothly changing trajectories is considered. In this case, since the velocity vector changes smoothly, this kind of ATS can be treated as LTS within a short time interval, and the velocity estimation method of LTS can be used in every short time interval. In our approach, only two appropriate spatial frequency points are selected by the established evaluation function in combining with a threshold value[7]. In addition to that the 1-D signals in selected points are not sensitive to the noise and background in the image sequence, the period of the 1-D signal is also limited to more than one period in the selection process, therefore the corresponding 1-D spectral estimation can be done by DFT using FFT algorithm.

This paper is organized as follows. The general moving object velocity estimation method of LTS and ATS in the MixeD is analyzed in Sect.2, then, the proposed method is described in Sect.3. Finally, simulation results and conclusions are presented in Sect.4 and Sect.5 respectively.

2. VELOCITY ESTIMATION IN THE MIXED

2.1. Representation of 3-D signal

Let $\mathbf{n} = (n_x, n_y, n_t)$ be a point in the discrete 3-D space N^3, then the 3-D discrete signal is defined at $\mathbf{R}$,

$$\mathbf{R} = \{\mathbf{n} | n_x, n_y = 0,1,\cdots,L-1,\ n_t = 0,1,\cdots,T_f-1\}$$

where n_x and n_y are spatial variables with total number L, n_t is the temporal variable with total frames T_f.

Let $\mathbf{D}_{mo}$ be the partial 3-D domain occupied by the signal of moving objects $s_{mo}(\mathbf{n})$, $b(\mathbf{n})$ be the signal of the background, then the composed image of moving objects and background can be expressed as follows,

$$x(\mathbf{n}) = s_{mo}(\mathbf{n}) + \hat{b}(\mathbf{n}) \tag{1}$$

where

$$\hat{b}(\mathbf{n}) = \begin{cases} 0 & \mathbf{n} \in \mathbf{D}_{mo} \\ b(\mathbf{n}) & \mathbf{n} \notin \mathbf{D}_{mo} \end{cases} \tag{2}$$

represents the background signal of a zero graylevel area changing along with time.

The actual observed 3-D signal is given by Eq.(3),

$$y(\mathbf{n}) = x(\mathbf{n}) + w(\mathbf{n}) \tag{3}$$

where $w(\mathbf{n})$ is the 3-D white noise.

The problem is to estimate the velocity (v_x, v_y) of moving object from $y(\mathbf{n})$ $(\mathbf{n} \in \mathbf{R})$.

2.2. Velocity Estimation Theory

To simplify the analysis, the image without background $b(\mathbf{n})$ and noise $w(\mathbf{n})$ is used. In this case, the linear trajectory signal (LTS) of moving object in an image sequence is defined as Eq.(4),

$$s_{LTS}(\mathbf{n}) = s_{2D}(n_x - v_x \cdot n_t, n_y - v_y \cdot n_t) \tag{4}$$

where $s_{2D}(n_x, n_y) = s_{LTS}(n_x, n_y, 0)$ is the signal of moving object at the initial frame $(n_t = 0)$, and (v_x, v_y) is the velocity vector of LTS.

For image sequence with two moving objects, the signal of moving objects can be written as the sum of two LTSs.

$$s_{mo}(\mathbf{n}) = s^1_{LTS}(\mathbf{n}) + s^2_{LTS}(\mathbf{n}) \tag{5}$$

In the MixeD, 2-D DFT is first applied to spatial variables n_x and n_y in Eq.(5), by using of Eq.(4), the signal is given as follows,

$$\begin{aligned} S_{mo}(\omega_x, \omega_y, n_t) &= DFT_{2D}\{s_{mo}(\mathbf{n})\} \\ &= \sum_{k=1}^{2} S^k_{LTS}(\omega_x, \omega_y) e^{-j(v^k_x \omega_x + v^k_y \omega_y) n_t} \end{aligned} \tag{6}$$

where the upper index k indicates two moving objects, and $S^k_{LTS}(\omega_x, \omega_y)$ is the corresponding 2-D DFT over spatial variables at $n_t = 0$.

For the object moving with a varying velocity vector in an image sequence without background, it can be defined as Eq.(7) similar to the LTS,

$$\begin{aligned} s_{mo}(\mathbf{n}) &= s_{ATS}(\mathbf{n}) \\ &= s_{2D}(n_x - v_x(n_t) \cdot n_t, n_y - v_y(n_t) \cdot n_t) \end{aligned} \tag{7}$$

where $(v_x(n_t), v_y(n_t))$ is the velocity vector of ATS.

Suppose the velocity vector changes smoothly, then in the short time interval, it can be treated as LTS. Similar to Eq.(6), the signal of two ATSs in the MixeD is given by

$$\begin{aligned} & S_{mo}(\omega_x, \omega_y, n_t) \\ & \approx \sum_{k=1}^{2} S^k_{ATS}(\omega_x, \omega_y) e^{-j(v^k_x(n_t)\omega_x + v^k_y(n_t)\omega_y) n_t} \end{aligned} \tag{8}$$

The varying temporal angular-frequencies $\omega^k_{ti}(n_t)$ $(k = 1, 2)$ contained in $S_{mo}(\omega_x, \omega_y, n_t)$ at the spatial frequency $(\omega_{xi}, \omega_{yi})$ is then given by

$$\omega^k_{t_i}(n_t) = -(v^k_x(n_t)\omega_{x_i} + v^k_y(n_t)\omega_{y_i}) \quad (k = 1, 2) \tag{9}$$

Therefore, the velocity estimation can be accomplished by solving the coupled equations of Eq.(9).

3. PROPOSED METHOD

In the MixeD, the temporal signals at spatial frequency point $(\omega_{xi}, \omega_{yi})(i = 1, 2, \cdots, M)$ are given by

$$Y_k(n_t) = DFT_{2D}\{y(\mathbf{n})\}|_{(\omega_x, \omega_y) = (\omega_{xi}, \omega_{yi})} \tag{10}$$

where M is the total number of spatial frequency points.

The proposition is to determine two spatial frequency points $(W_{xi}, W_{yi})(i=1,2)$ through the evaluation of the 1-D spectrum $|DFT\{Y_k(n_t)\}|$, and estimate the corresponding two peak frequencies $W_{ti}^k(k=1,2)$. Then substitute ω_{ti}^k and $(\omega_{xi}, \omega_{yi})$ with W_{ti}^k and (W_{xi}, W_{yi}) respectively in Eq.(9).

3.1. Spatial Frequency Selection Method

The goal of spatial frequency selection is to search for the spatial frequency point where the 1-D complex signal is less influenced by the background and noise. For the image sequence with objects having two motion vectors, the 1-D power spectrum $|DFT\{Y_k(n_t)\}|$ should have two main peaks theoretically, then the selection can be accomplished by evaluating the 1-D power spectrum[7]. The evaluation function corresponding to the image sequence with two motion vectors is established as

$$J_2 = \lambda_2 \frac{(H_{p1} + H_{p2})}{|H_{p1} - H_{p2}|} + |P_{p1} - P_{p2}| \tag{11}$$

where λ_2 is the weighting factor, P_{p1}, P_{p2} and H_{p1}, H_{p2} are the peak frequencies and the peak heights of the power spectrum as shown in Fig.1.

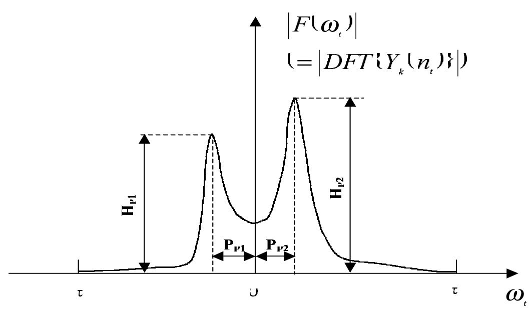


Fig. 1: Evaluation of 1-D power spectrum

For a given number of frames N_f used in spectral estimations, the signals should be with at least more than one period in order to achieve a precise spectral estimation, it is guaranteed by the following inequality

$$2\pi/\omega_t(n_t) \leq N_f \tag{12}$$

To determine the first appropriate spatial frequency point (W_{x1}, W_{y1}), we seek the point with maximum evaluation function value J_2 under the limitation of Eq.(12). To select the second point (W_{x2}, W_{y2}), in addition to the same selection as (W_{x1}, W_{y1}), a threshold Th given by Eq.(13) should also be satisfied to guarantee the independence of the coupled equations of Eq.(9).

$$Th \leq |W_{x_1} W_{y_2} - W_{y_1} W_{x_2}| \tag{13}$$

3.2. Background Pre-Processing

To make the velocity estimation more precisely, it is better to process the background in advance, this is accomplished by two kinds of background pre-processing methods. The first background pre-processing method (BPM1) is the image thresholding based on image segmentation technique. It is accomplished by selecting a graylevel threshold value according to the histogram of the image, and subtracting the image with this threshold value. This method can attenuate the influence of the static background in some extent.

The second pre-processing method (BPM2) is the removal of the DC component from the 1-D signal. This is accomplished by subtracting the average component from the 1-D signal.

3.3. Velocity Estimation

Since the moving objects have varying velocity vectors, it is better to select the appropriate spatial frequency points adaptively. This is achieved by using the spatial frequency selection method in each estimation interval $(n_t-(N_f-1)/2, n_t+(N_f-1)/2)$ respectively. Let $f_p(n_t)$ be the processed 1-D complex signal in the MixeD, the spatial frequency points selection is carried every N_f frames as shown in Fig.2.

In practice, the image sequence is pre-processed by the BPM1 firstly, then the problem is transformed into the mixed domain, and the obtained 1-D complex signal is processed by the BPM2. The velocity estimation is then carried over the adaptively selected spatial frequency points. The velocity estimation process in each estimation interval is shown in Fig.3.

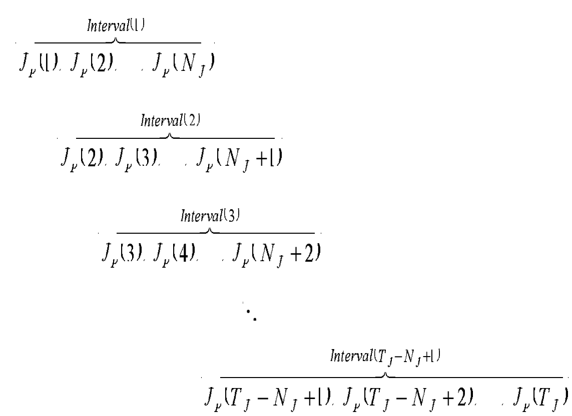


Fig. 2: Velocity estimation interval

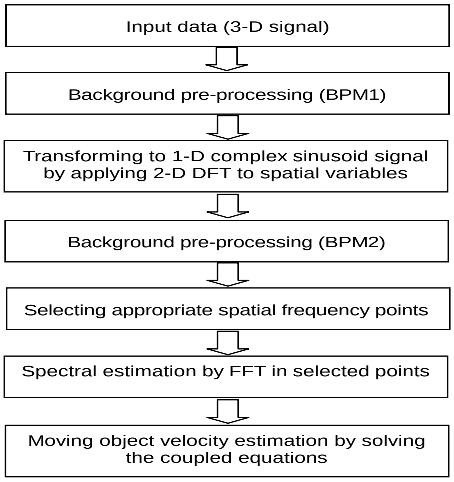


Fig. 3: Velocity estimation process

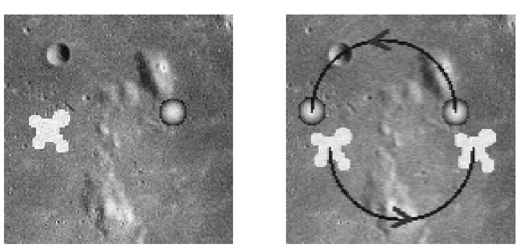
4. SIMULATION RESULTS

In the simulation, the image sequence is with background and the white noise of zero mean and standard deviation $\sigma = 0.05$ is added. The image size is 128×128, the total frames are 200 frames, the intensity is $[0, 1]$, and the interval used in velocity estimation is $N_f = 31$ frames.

The first moving object I is with uniform intensity 1, and has a complicate shape as shown in Fig.4. The second moving object II is a disc ($radius = 8$) of non-uniform intensity with the highest intensity 1 in the center. The image at the initial frame and the sketch of its central trajectory is shown in Fig.4(a) and (b) respectively. The ideal velocity vectors in pixels/frame (ppf) for both objects are as follows,

$$(v_{x1}, v_{y1}) = (\frac{\pi}{5} sin(\frac{t-1}{200}\pi), \frac{\pi}{5} cos(\frac{t-1}{200}\pi))\ ppf \quad (14)$$

$$(v_{x2}, v_{y2}) = (-\frac{\pi}{5} sin(\frac{t-1}{200}\pi), -\frac{\pi}{5} cos(\frac{t-1}{200}\pi))\ ppf \quad (15)$$



(a) (b)

Fig. 4: Initial image and the trajectory sketch

The velocity vector estimated every $5frame$ for moving object I is given in Fig.5, while the velocity vector estimated every $5frame$ for moving object II is shown in Fig.6. It is obvious that both motion vectors have been correctly estimated.

5. CONCLUSION

In this paper, the motion vectors of two moving objects with arbitrary trajectory have been correctly estimated in the MixeD. Since the velocity estimation is carried over two appropriate spatial frequency points (W_{x1}, W_{y1}) and (W_{x2}, W_{y2}) selected in the 2-D spatial frequency domain, a good result has been acquired even though the image sequence is with background and noise. Additionally, the spectra over the selected points are estimated by FFT spectral estimation, this makes the computation time shorter.

REFERENCES

[1] L.T.Bruton,N.R.Bartley, Three-Dimensional Image Processing Using the Concept of Network Resonance, IEEE Trans. on Circuits and System, Vol. CAS-32, No.7, July 1985.

[2] D.J.Heeger, Optical flow from spatialtemporal filters, IEEE 1th Int. Conf. on Computer Vision, London, pp.181-190, June 1987.

[3] K.S.Knudsen,L.T.Bruton, Moving Object Detection and Trajectory Estimation in the Transform/Spatiotemporal Mixed Domain, IEEE Int. Cont on Acoustics Speech and Signal Processing, pp.505-508, March 1992.

[4] K.S.Knudsen,L.T.Bruton, Moving Object Nonlinear Trajectory Estimation in the Transform/Spatiotemporal Mixed Domain, IEEE Int. Cont. Symposium on Circuits and Systems, Vol.5, pp.2481-2483, 1992.

[5] I.Nakagawa, T.Tago, K.Kondo, N.Hamada, Trajectroy Estimation and Detection of Moving Objects in the Transform/Spatio-temporal Mixed Domain, ACCV'95, pp.712-716 1995.

[6] S.A.Rajala,A.N.Riddle,W.E.Snyder, Application of the One-Dimensional Fouier Transform for Tracking Moving Objects in Noisy Environments, Computer Vision,Graphics and Image processing, Vol.21, pp.280-293, 1983.

[7] S.L.Wu, N.Hamada, An Appropriate Spatial Frequency Selection Method for Moving Object Velocity Estimation in the Mixed Domain, submitted to IEICE Trans. on Fundamentals.

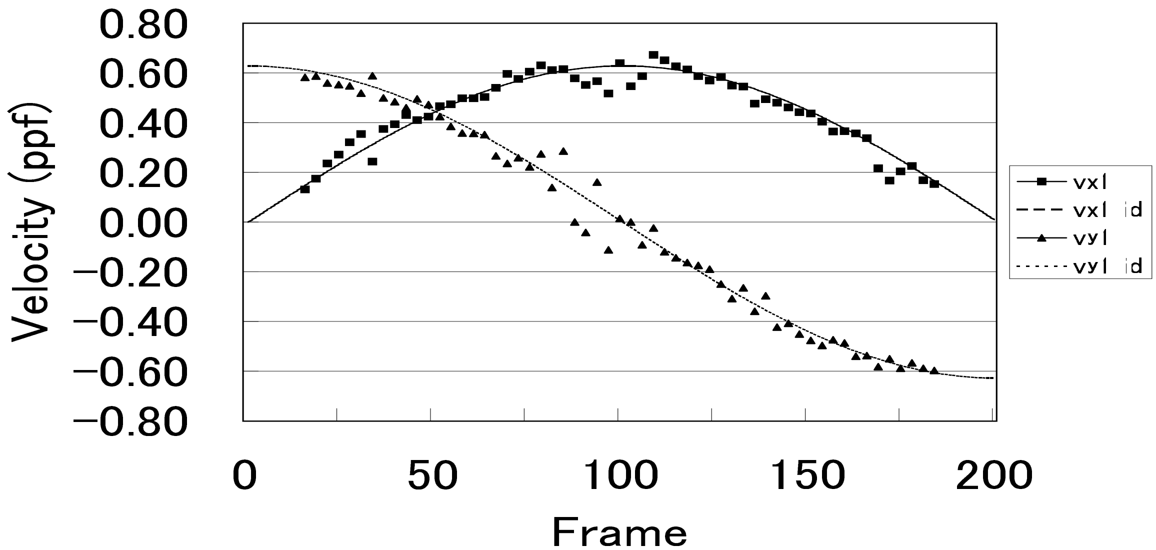


Fig. 5: Velocity estimated for object I

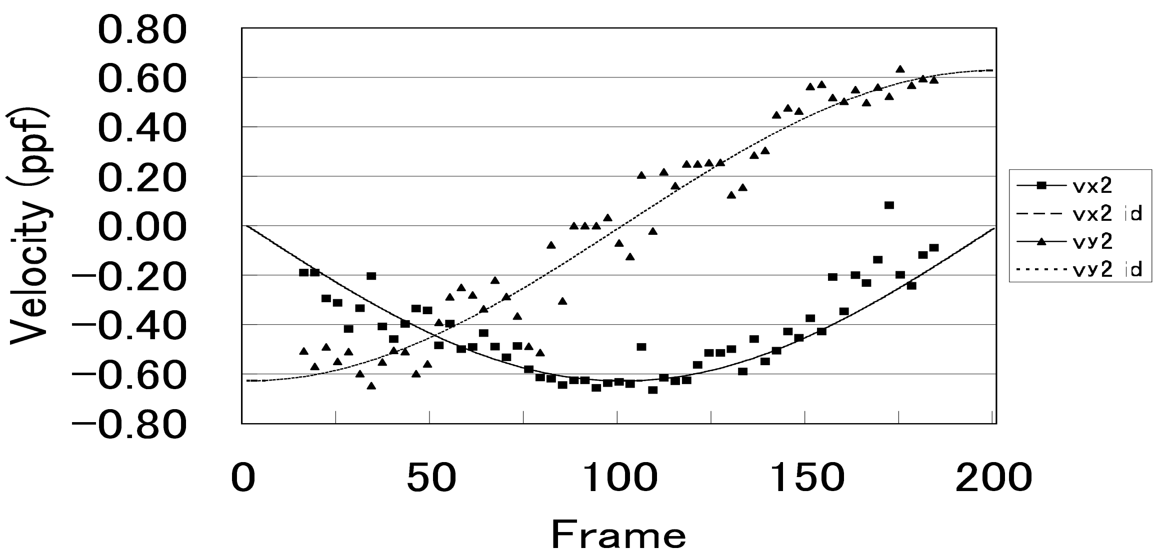


Fig. 6: Velocity estimated for object II