Wiener Filter for Zoom Parameter Estimation in Three-Parameter Motion Model

Pei-Chuan Liu*, Wen-Zen Shen and Wen-Thong Chang

National Chiao Tung University, Department of Electronics Engineering, Hsin-Chu 30039, Taiwan, R.O.C

ABSTRACT

From the linearized Taylor series expansion, an iterative, gradient-based method is used to estimate the zoom and pan motion parameters. In order to take into account the high order expansion error, the Wiener filtering techniques are investigated. It is shown that the expansion error can be efficiently removed with the Wiener filter. However, the reliability of the estimated parameters is still highly dependent upon the accuracy of the gradient information of the images.

keyword : Motion parameter estimation, Wiener filter, zooming factor estimation

1. INTRODUCTION

Motion parameters describe the relative motion phenomenon between image objects in the image sequence. Some cases of motions are due to the zoom and pan of the camera. The 2-parameter motion model used for video coding assumes only 2-D translatory image motion. However, in many cases, translatory motion assumption is not sufficient to describe the image motion phenomenon. For this, a three-parameter motion model has been proposed.1 In the proposed,' one parameter is used to describe the ratio of the focal lengthes before and after zooming and the other two parameters are used to describe the pan of the camera.

The criterion for most motion estimation algorithms³⁻⁷ is to minimize the displaced frame difference. For this, a signal representation model is used to describe the luminance signal as a function of the motion parameters based on the Taylor series expansion. In this signal representation, the luminance of the estimated corresponding point is expressed as the sum of the luminance of the true corresponding point and a function of the image gradient and the parameter estimation error. This parameter estimation error is defined as the difference between the true parameter value and the current estimated one. In the proposed,¹ a direct pseudo-inverse method is used to calculate the motion parameters from the observed displaced frame difference and the image gradient. However, the performance in the proposed' is not very satisfactory. Large variation in the search process is commonly seen and sometimes the algorithm diverges. The problem with the direct pseudo-inverse method is due to the high order expansion terms in the Taylor series expansion of the luminance signal. These high order expansion terms are not small initially and can not be neglected. This can be due to the poor estimate of the gradient or that the parameter error is too large initially. To deal with the high order expansion terms, in this paper, a Wiener-based recursive motion estimation algorithm is proposed. With Wiener based algorithm, the high order expansion terms are treated as a random process. With such random data model, the displaced frame difference is modelled as a function of the image gradient and the parameter estimation error plus a high order expansion error. The Wienerbased algorithm is a linear estimator that derives the motion parameters by minimizing the mean square difference between the estimated parameter error and the true error. Due to the minimization of the mean square difference, second order statistics of the parameter estimation error and the random process that models the high order expansion error are needed. With the consideration of such high order expansion error, the convergence and the stability of the Wiener-based algorithm is far superior to that of the direct pseudo-inverse method.

Wiener algorithm has been applied to estimate the 2-dimensional displacement⁵⁻⁷. In the twoparameter model, the parameters are assumed to be uncorrelated and a constant diagonal correlation matrix is used. However, the zoom and pan discussed in this paper are correlated. Therefore, the correlation matrix of the estimation error is significant. A procedure has been proposed to estimate such correlation matrix. If the motion parameter error is within a certain range, the high order expansion error can be assumed to be uncorrelated for each image point. Therefore the corresponding correlation matrix is assumed to be a diagonal matrix.

In the following section, we first briefly review the three-parameter motion model. Then the signal representation model that describes the relationship between the motion parameters and the observed displaced frame difference is derived. In section 3, the Wiener-based algorithm for parameter estimation as well as the estimation of the correlation matrices are analyzed. The simulation results are shown in section 4 and a conclusion is made in section 5.

2. THE SIGNAL REPRESENTATION MODEL

The three-parameter motion model can be described as the transformation of the coordinate

$$
(X_2, Y_2) = A(X_1, Y_1) = (a_1 X_1 + a_2, a_1 Y_1 + a_3)
$$
\n(1)

where (X_1,Y_1) and (X_2,Y_2) are the coordinates of an image point before and after zoom and pan of the camera. Parameters a_1, a_2 and a_3 are the three motion parameters to be estimated. When parameter $a_1 = 1$, the three-parameter model will degenerate to the traditional translational two-parameter motion model with $a_2 = dx$ and $a_3 = dy$.

Based on the motion model, a signal representation model is used to describe the relationship between the luminance signal and the motion parameters. Form the signal representation model, Wiener algorithm is derived to estimate the motion parameters from the luminance signal. Let the luminance at position (x, y) in frame k be defined as $S_k(x, y)$. If the motion model is assumed to be $A(x, y)$ y) = $(a_1 x + a_2, a_1 y + a_3)$, then the relationship between $S_k(x, y)$ and $S_{k+1}(x, y)$ can be described as

$$
S_{k+1}(x, y) = S_k(A(x, y))
$$
\n⁽²⁾

The displaced frame difference(DFD) is defined as

$$
DFD(x,y) = S_{k+1}(x,y) - S_k(\widehat{A}(x,y)) = S_{k+1}(x,y) - S_k(\widehat{a}_1 x + \widehat{a}_2, \widehat{a}_1 y + \widehat{a}_3)
$$
(3)

where a_1, a_2, a_3 are the current estimates of the true motion parameters a_1, a_2 and a_3 .

If a_1, a_2 , and a_3 approachs to a_1, a_2 and a_3 respectively, then DFD will approach to zero. Since $S_{k+1}(x,y) = S_k(A(x,y))$, the DFD can be expressed as $S_k(A(x,y))$ - $S_k(A(x,y))$. That is the DFD is actually the result of the error of the estimate of the motion parameters. Using the Taylor series expansion, the luminance function at point $p_o = A(x, y)$ with estimated $e = (a_1, a_2, a_3)$ can be evaluated with respect to the true motion parameter a_i as

$$
S_k(A(x, y)) = S_k(\widehat{A}(x, y)) + \sum_{i=1}^3 \frac{\partial S_k}{\partial a_i} \Big|_{p_0, e^{(a_i - \widehat{a}_i)} + r(x, y)}
$$
(4)

where $r(x, y)$ denotes the high order expansion terms in this Taylor series expansion.

Rewrite Eqn.(4) we have

$$
DFD(x,y) = G_{a1}(x,y) (a_1 - \hat{a}_1) + G_{a2}(x,y) (a_2 - \hat{a}_2) + G_{a3}(x,y) (a_3 - \hat{a}_3) + r(x,y) \nwhere G_{a_i}(x,y) = \frac{\partial S_k}{\partial a_i} |_{p_o,e}
$$
\n(5)

591

In order to describe Eqn.(5) directly from the image signal, the gradients with respect to the motion parameters can be replaced by the gradients in X and Y coordinates. By defining $(x', y') = A(x,y)$, we have

$$
\frac{\partial S_k}{\partial a_i} \Big|_{p_o,e} = \frac{\partial S_k}{\partial x'} \frac{\partial x'}{\partial a_i} \Big|_{p_o,e} + \frac{\partial S_k}{\partial y'} \frac{\partial y'}{\partial a_i} \Big|_{p_o,e}
$$
\n
$$
\frac{\partial x'}{\partial a_1} = x \qquad \frac{\partial x'}{\partial a_2} = 1 \qquad \frac{\partial x'}{\partial a_3} = 0
$$
\n
$$
\frac{\partial y'}{\partial a_1} = y \qquad \frac{\partial y'}{\partial a_2} = 0 \qquad \frac{\partial y'}{\partial a_3} = 1
$$
\n(7)

Thus,

$$
DFD(x,y) = S_{k+1}(x,y) - S_k(\stackrel{\frown}{A}(x,y))
$$

= $S_k(A(x,y)) - S_k(\stackrel{\frown}{A}(x,y))$
= $(G_x x + G_y y) \times (a_1 - \stackrel{\frown}{a}_1) + G_x \times (a_2 - \stackrel{\frown}{a}_2) + G_y \times (a_3 - \stackrel{\frown}{a}_3) + r(x,y)$
= $(G_x x + G_y y) u_1 + G_x u_2 + G_y u_3 + r(x,y)$ (8)

where $G_x = \frac{\partial S_k}{\partial x'}\big|_{p_0, e}$ and $G_y = \frac{\partial S_k}{\partial y'}\big|_{p_0, e}$ are the X and Y directional gradient functions at

the point $A(x', y')$ and u_i is the estimation error defined as the difference between a_i and $\hat{a_i}$.

Eqn.(12) describes the DFD as a function of the parameter error u_1 , u_2 and u_3 . The expansion error can be seen as the error in the estimation of the DFD. As can be seen from Eqn.(8), the product term $(G_x x + G_y) u_1 + G_x u_2 + G_y u_3$ can be seen as an estimate of the DFD value.

From Eqn.(8), it is found that the three-parameter signal representation model is similar to the two-parameter model. As derived in the proposed⁵, the two-parameter signal representation model is

$$
DFD(x, y) = G_x u_x + G_y u_y + r'(x, y)
$$
\n(9)

with $u_x = dx - dx$ and $u_y = dy - dy$ as the X and Y directional estimation error. Since $u_x = u_1 x + u_2$ and $u_y = u_1 y + u_3$, It can be seen that Eqn.(8) and Eqn.(9) are similar. The difference between Eqn.(8) and $Eqn(9)$ are the description of the motion phenomenon. The underlying search procedure for point correspondence is the same. Both equations indicate the discrepancy in point correspondence as the product of gradient and parameter errors. Therefore, the characteristics of the expansion error $r(x,y)$ and $r'(x,y)$ are the same. Based on this fact, those features and assumptions that have been proposed⁶ can still be used in the estimation of the zoom and pan.

For N points of signals, Eqn. (8) can be written in a matrix form as

$$
\mathbf{D} = \mathbf{G}(\mathbf{A} - \mathbf{A}^{(p)}) + \mathbf{R}
$$

= $\mathbf{G}\mathbf{u} + \mathbf{R}$ (10)

with

where **with**

 $A = [a_1, a_2, a_3]^T$ is the true motion parameter vector. $A^{(p)} = [\hat{a}_1^{(p)}, \hat{a}_2^{(p)}, \hat{a}_3^{(p)}]^T$ is the estimated motion parameter vector at iteration p. $\mathbf{u} = [u_1, u_2, u_3]^{T}$ is the estimation error vector.

From Eqn. (10), it can be seen that the estimation error \bf{u} between the true value and the estimated one is embedded in the displaced frame difference and is not explicity calculatable due to the noise process R.

3. THE WIENER-BASED ESTIMATION METHOD

If the expansion error is neglected, the estimation becomes a simple pseudo-inverse computation1. However, to obtain better estimation, this random error should be considered. To recover signal from random noise, it is known that the Wiener filter is a very effective linear estimator. From the signal model, the desired signal is Gu and the observed signal is DFD. The purpose is to recover Gu form DFD. Therefore, in the following, we first derive the Wiener fillering process to show that Gu can indeed be recovered from DFD. Since G is known u can be easily calculated. These two processes can be combined into one for simplicity.

The Wiener-based algorithm is derived based on the criterion that the mean square value $E\{\parallel u \mathbf{u}^{(p)}$ \parallel^2 } is minimized. Let the linear estimator be denoted as L. The input and the output relation of the linear estimator can be described as

$$
\mathbf{u}^{(p)} = [u_1^{(p)}, u_2^{(p)}, u_3^{(p)}]^T = \mathbf{\widehat{A}}^{(p+1)} - \mathbf{\widehat{A}}^{(p)} = \mathbf{L} \mathbf{D}
$$
\n(11)

According to the Wiener based process⁵⁻⁶, the linear estimator can be described

$$
\mathbf{L} = (\mathbf{P}_{\mathbf{u}} \mathbf{G}^T + \mathbf{P}_{\mathbf{u}} \mathbf{R}) (\mathbf{G} \mathbf{P}_{\mathbf{u}} \mathbf{G}^T + \mathbf{P}_{\mathbf{R}} + \mathbf{G} \mathbf{P}_{\mathbf{u}} \mathbf{R}^+ \mathbf{P}_{\mathbf{u}} \mathbf{R}^T \mathbf{G}^T)^{-1} \tag{12}
$$

where $P_u = E\{uu^T\}$, $P_R = E\{RR^T\}$ and $P_{uR} = E\{uR^T\}$ is the correlation matrix of u and R. Eqn.(12) is the Wiener based estimation algorithm for 3-parameter motion model. In this algorithm, G is assumed to be a known constant. From Eqn.(12), it can be seen that the purpose of the Wiener filter is to use the knowledge of P_u , P_R and P_u _R to estimate Gu form DFD. Its performance is entirely dominated by the knowledge of these three matrices. Therefore, to discuss the performance of the Wiener filter, we need to first discuss the estimation of these three matrices.

A. the estimation of P_{uR}

As discussed above, the characteristics of the expansion errors in the signal representation model for the 2-parameter and the 3-parameter cases are similar. For the 2-parameter model, the assumption that u is uncorrelated with r has been made based on the observation that the expansion error is a random signal with zero mean. For the 3-parameter model, as derived in Section II, the expansion error is still a random signal with zero mean when the estimation error is small. Therefore, the assumption that the random signal r is uncorrelated with the signal u is still made in this 3parameter model. With $E\{\mathbf{u}\mathbf{R}^T\} = E\{\mathbf{R}\mathbf{u}^T\}^T = 0$, Eqn.(12) will be

$$
\mathbf{L} = \mathbf{P}_{\mathbf{u}} \mathbf{G}^T (\mathbf{G} \mathbf{P}_{\mathbf{u}} \mathbf{G}^T + \mathbf{P}_{\mathbf{R}})^{-1}
$$

From the matrix manipulation in the proposed⁵, we have

$$
\mathbf{L} = (\mathbf{G}^T \mathbf{P}_{\mathbf{R}}^{-1} \mathbf{G} + \mathbf{P}_{\mathbf{u}}^{-1})^{-1} \mathbf{G}^T \mathbf{P}_{\mathbf{R}}^{-1}
$$
(13)

and the Wiener based algorithm becomes

$$
\mathbf{A}^{(p+1)} - \mathbf{A}^{(p)} = \mathbf{u}^{(p)} = \mathbf{L} \mathbf{D} = (\mathbf{G}^T \mathbf{P}_{\mathbf{R}}^{-1} \mathbf{G} + \mathbf{P}_{\mathbf{u}}^{-1})^{-1} \mathbf{G}^T \mathbf{P}_{\mathbf{R}}^{-1} \mathbf{D}
$$
\n(14)

B the estimation of $P_u = E{uu}^T$

For the 2-parameter motion model, in the proposed⁵, the estimation error u_x and u_y are assumed to be zero mean processes and are uncorrelated to each other. Thus, the correlation matrix P_u is assumed to be a diagonal identity matrix scaled by the constant variance σ_u^2 . But in 3-parameter motion model, there exist a strong relationship among the parameters a_1 , a_2 and a_3 . In this paper the matrix P_u is estimated as

$$
\mathbf{P}_{\mathbf{u}}^{(p+1)} = \frac{p}{p+1} \mathbf{P}_{\mathbf{u}}^{(p)} + \frac{1}{p+1} \mathbf{u}^{(p+1)T} \mathbf{u}^{(p+1)}
$$
(15)

C the estimation of $P_R = E\{RR^T\}$

From the second order Taylor series expansion, the variance of the expansion error σ_r^2 for each image point is estimated as

$$
\sigma_{n}^{2} = (G_{xy})^{2} + \frac{1}{2} G_{xx} G_{yyi}) \sigma_{u}^{4}
$$
 (16)

The correlation matrix P_R is assumed to be diagonal with each element as σ_{ri}^2 .

4. ThE PERFORMANCE EVALUTION

The Wiener based algorithm is derived from the signal representation model based on the first order Taylor series expansion. The first order Taylor series expansion is feasible only when the change of intensity is smooth. The convergence of the Wiener based algorithm is also largely dependent on the accuracy of the gradient estimation. With Wiener algorithm, the purpose is to filter out the expansion error. The estimation result is the product term Gu. With the knowledge of the gradient G, u is estimated to find the new corresponding points for subsequent iteration. If the knowledge of G is poor, u will not be correct. Then the estimation will take longer time or even diverge. Due to the characteristics of the motion model, the coordinates of new corresponding points are usually not integer and are not available. For this, the bilinear interpolation is used to find the luminance of these new corresponding points.

To show that the Wiener filter can indeed recover Gu from the noise corrupted DFD, two synthetic images shown in Fig.1 consisting of gaussian distributions are used to demonstrate the effectiveness of the algorithm. The first image contains two gaussian distributions $S_{11}(x, y)$ =255

 $\exp(\frac{-(x^2 + (y+30)^2)}{200})$, $S_{12}(x, y) = 255 \exp(\frac{-(x^2 + (y-30)^2)}{200})$. The second image contains $S_{21}(x, y) = 255$ $exp(\frac{((x)^2 + (y+50)^2)}{200 \times (1.00)^2})$, $S_{22}(x, y)$ =255 $exp(\frac{((x-1)^2 + (y-51)^2)}{200 \times (1.00)^2})$. Form these two images, two motion $200\times(1.08)^2$ 200 $\times(1.08)^2$

parameters $A_1 = (1.08, 0, 0)^T$ and $A_2 = (1.08, 1, 1)^T$ for the two distributions are observed. In this simulation, the gradients are estimated with three different methods :

- 1. The optimum method :
 $G_x(x_i, y_i) = 255 \exp(-(x_i^2 + y_i^2)/(200(1.08)^2)) \times (-2x_i/200(1.08)^2)$ $G_y(x_i, y_i) = 255 \exp(-(x_i^2+y_i^2)/(200(1.08)^2)) \times (-2y_i/200(1.08)^2)$
- 2. The 6-point estimation

$$
G_x(x_i, y_i) = \frac{1}{4} \left\{ S_k(x_i+1, y_i-1) - S_k(x_i-1, y_i-1) \right\}
$$

+ $\frac{1}{2} \left\{ S_k(x_i+1, y_i) - S_k(x_i-1, y_i) \right\}$
+ $\frac{1}{4} \left\{ S_k(x_i+1, y_i+1) - S_k(x_i-1, y_i+1) \right\}$
 $G_y(x_i, y_i) = \frac{1}{4} \left\{ S_k(x_i-1, y_i+1) - S_k(x_i-1, y_i-1) \right\}$
+ $\frac{1}{2} \left\{ S_k(x_i, y_i+1) - S_k(x_i, y_i-1) \right\}$
+ $\frac{1}{4} \left\{ S_k(x_i+1, y_i+1) - S_k(x_i+1, y_i-1) \right\}$

3. The 2-point estimation

 $G_x(x_i, y_i) = S_k(x_i+1, y_i) - S_k(x_i-1, y_i)$
 $G_y(x_i, y_i) = S_k(x_i, y_i+1) - S_k(x_i, y_i-1)$

The simulation results for these two gaussian distributions are shown in Fig.2 and Fig.3. Fig.2 shows the estimation of A corresponding to the movement of S_{11} to S_{21} . Fig.3 shows the estimation of A corresponding to the movement of S_{12} to S_{22} . Both cases show that the Wiener filter can indeed recover $\tilde{G}u$ from the signal representation model. However, as can be seen from both figures, the effect of the gradients is very significant. In the first case, with only zooming motion the effect of poor gradient is only on the convergence speed. For the second case, with simultaneous zooming and translation, the accuracy of the motion parameters is affected. This indicates the importance of the gradient estimation.

5. CONCLUSION

A signal representation model based on the first order Taylor expansion is used to describe the relation among motion parameters, gradient and DFD. Due to the use of Taylor expansion, high order . expansion terms are treated as random noise in the signal representation model. To recover the motion parameters with such model, the Wiener-based algorithm is shown to be very effective in dealing with the random expansion error. The problem of applying such algorithm to the real images for motion parameter estimation is the difficulty of good gradient estimation. Generally, in a smooth region with reliable gradient estimation, the algorithm can derive the desired solution.

REFERENCE

[1] M.Hotter, "Differential Estimation of The Global Motion Parameters Zoom and Pan", Signal Processing, vol.16, no.3, March 1989, pp.249-265.

- [2] Y. Nakaya and H. Harashima, "Motion Compensation Based on Spatial Transformations", IEEE Trans. on CAS for Video Technology, vol.4, no.3, Jun. 1994, pp-339-356.
- [3] Walker, D. R. and K. R. Rao, "Improved Pel-Recurssive Motion Compensation", IEEE Transactions on Commucations, vol. COM-32, no. 10, December 1984, pp.1128-1134.
- [4] C. Cafforio and F. Rocca, "The Differential Method for Motion Estimation", in Image Sequance Processing and Dynamic Scene Analysis (Nato ASI series, vol.12).T. S. Huang, Ed. New York: Springer-Verlag, 1983, pp.104-124.
- [5] Biemond, j. , L. Looijenga, D. E. Boekee and R. H. J. M. Plompen, "A Pel-Recursive Wiener-Based Displacement Estimation Algorithm", Signal Processing, vol. 13, December 1987, pp.399-412.
- [6] L. Boroczky, j. N. Driessen and 3. Biemond, "Adaptive Algorithms for Pel-Recursive Displacement Estimation", SPIE Vissual comm. and Image Proc., vol.1360, 1990, pp.1210-1221.
- [7] 3. Biemond, J.N. Driessen, A.M. Geurtz and D.E. Boekee, "A pel-recursive Wiener-based algorithm for the simutaneous estimation of rotation and translation", SPIE Vissual comm. and Image Proc., vol.1001, 1988, pp.917-924.

 $\sf(a)$

(b)

Fig.1 Two synthetic images consisting of gaussian distributions (a) image consists two gaussian distributions $S_{II}(x, y) = 255 \exp(\frac{-(x^{2}+y^{2}+y^{2})}{200})$, $S_{II}(x, y) = 255$ $exp(\frac{-(x^{2}+y^{2}-50)}{200})$. (b) image contains $S_{21}(x, y) = 255 \exp(\frac{-(x^{2}+y^{2}-50)}{200 \times (1.08)^{2}})$, $S_{22}(x, y)$ $=255 \exp(\frac{-(x-1)^2+(y-51)^2}{2})$ $200\times(1.08)^2$

(c)

Fig.2 The estimated motion parameter a_1 , a_2 and a_3 with three different gradient estimation methods for the Gaussian image. (the desired motion parameters (1.08,0, 0))

598

 (a)

 (c)

The estimated motion parameter a_1, a_2 and a_3 with three different gradient estimation $Fig.3$ methods for the Gaussian image. (the desired motion parameters (1.08, 1, 1))

 \bar{z}