# THE EFFECT OF GLOBAL MOTION PARAMETER ACCURACIES ON THE EFFICIENCY OF VIDEO CODING

*Gökçe Dane and Truong Q. Nguyen*

Department of Electrical and Computer Engineering
University of California at San Diego, 9500 Gilman Drive, La Jolla, CA, 92093-0407
gdane@ucsd.edu, nguyent@ece.ucsd.edu

## ABSTRACT

In this paper, we present theoretical analysis on how the global motion parameter accuracies affect the efficiency of motion compensated video coding. The inaccurate global motion compensation is modelled by introducing probabilistic rotation, scale and translation parameter errors. Approximate expressions that relate the power spectrum of the prediction error to motion parameter errors is derived. By doubling the accuracy of the motion parameters, up to 6dB theoretical gains can be obtained in prediction error variance.

## 1. INTRODUCTION

Global motion estimation (GME) is essential in several applications such as video compression, segmentation, mosaicing, and image registration. Global motion refers to higher order motion models like affine or bilinear, which improves the prediction gain over simple translational model.

Although many global motion estimation techniques have been proposed, the parameter accuracy vs coding rate analysis have not been performed. Efficiency of motion compensated coding with translational motion errors have been studied in [1, 3]. Optimizing the translational accuracy for block-based coders is analyzed in [2].

This paper studies the effect of GM parameter inaccuracies on motion compensated video coding. The 4-RST (rotation, scale translation) motion model used in the analysis is described in Section II. The power spectrum of the prediction error is related to GM parameter errors, and approximate expressions for prediction error spectrum is derived in Section III. The effect of motion accuracy level, the effect of video signal model, and finally the effect of approximation on the prediction gain are studied in Section IV, which is followed by conclusion in the last section.

This work is supported by Skyworks Inc. and by UC DIMI.

## 2. MOTION MODEL

Let two images $I_1(x, y)$ and $I_2(x, y)$ be related to each other by an affine transform. The coordinates $\mathbf{v_2} = (x^{(2)}, y^{(2)})^T$ in $I_2$ can be obtained from the coordinates $\mathbf{v_1} = (x^{(1)}, y^{(1)})^T$ in $I_1$ by the following relation.

$$\mathbf{v_2} = \mathbf{A}\mathbf{v_1} + \mathbf{t}, \;\; \mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \;\text{ and }\; \mathbf{t} = \begin{pmatrix} t_x \\ t_y \end{pmatrix} \quad (1)$$

The 6 parameters $(a_{11}, a_{12}, a_{21}, a_{22}, t_x, t_y)$ describe the affine model completely. Setting $a_{11}$=$a_{22}$, and $a_{12}$=$-a_{21}$ will give the 4-RST model, which is commonly used in global motion estimation. Let $a$ and $\theta$ be the scale and rotation parameters respectively, then $a_{11} = a\cos\theta$, $a_{12} = -a\sin\theta$ with $\mathbf{t} = (t_x, t_y)^T$. In this paper, we analyze how these rotational, scale and translational errors affect the efficiency of video coding. The next section models the inaccurate global motion compensation and derives expressions for prediction error variance for each individual motion parameter.

## 3. PREDICTION ERROR ESTIMATION FOR GLOBAL MOTION COMPENSATION

This section develops the approximate expressions for the motion compensated prediction error as functions of motion parameter accuracies. Assume that $s(x, y)$ is the video signal that is predicted in the current frame, and $b(x, y)$ is the estimate of $s(x, y)$. The prediction error can be written as $e(x, y)$= $s(x, y)$-$b(x, y)$. If perfect motion vector parameters are used, it would be possible to obtain $s(x, y)$=$b(x, y)$ without any error. Unfortunately, we are bound to have errors in motion parameters. The motion parameters have to be transmitted with a limited bit rate. Therefore, there will be errors due to inaccuracy in these parameters; which will be represented as $\Delta a$, $\Delta\theta$, $\Delta_x$ and $\Delta_y$.

The prediction error variance $\sigma_e^2$ which is defined in eq.(2) will be used as the error measure,

$$\sigma_{pe}^2 = \frac{1}{4\pi^2}\int\int \Phi_{ee}(\omega_x, \omega_y)d\omega_x d\omega_y \quad (2)$$

0-7803-8554-3/04/$20.00 ©2004 IEEE.

where $\Phi_{ee}$ is the power spectral density of the error, defined as

$$\Phi_{ee}(\omega_x,\omega_y)=\mathcal{F}\{E[e(x,y),e(x,y)^*]\} \quad (3)$$

$\mathcal{F}$ represents the Fourier transform, and $E[\cdot]$ is the expectation operator.

### 3.1. Translational Error

Assume that the image only has translational motion $\mathbf{t}$=($mv_x$, $mv_y$). Let the estimate of this motion vector be $\mathbf{t_e}$=($mv_{ex}$, $mv_{ey}$). Then, error in translational motion parameter will be $\Delta_{\mathbf{t}}=(\Delta_x,\Delta_y)$. The motion compensated block $b(x,y)$ can be written in terms of translational error and original video signal in current frame as $b(x,y)=s(x-\Delta_x,y-\Delta_y)$. Substituting $b(x,y)$ in the prediction error equation yields

$$e(x,y)=s(x,y)-s(x-\Delta_x,y-\Delta_y) \quad (4)$$

Expanding $s(x-\Delta_x,y-\Delta_y)$ by Taylor's series around $(x,y)$ with the first approximation term and substituting it in eq.(4) yields

$$e(x,y)=\Delta_x s_x+\Delta_y s_y \quad (5)$$

where $s_x=\frac{\partial s}{\partial x}$ and $s_y=\frac{\partial s}{\partial y}$. Substituting eq.(5) in eq.(3) yields the following psd of the prediction error:

$$\Phi_{ee}(\omega_x,\omega_y)=\mathcal{F}\{E[(\Delta_x s_x+\Delta_y s_y)(\Delta_x s_x+\Delta_y s_y)^*]\} \quad (6)$$

We assume that translational motion error $\Delta_x$ and $\Delta_y$ are independent from each other, they both have zero mean ($E[\Delta_x]=E[\Delta_y]=0$) and equal variance ($E[\Delta_x^2]=E[\Delta_y^2]=E[\Delta^2]$). With these assumptions, eq.(6) is written as

$$\Phi_{ee}=E[\Delta_x^2]\mathcal{F}\{E[s_x s_x^*]\}+E[\Delta_y^2]\mathcal{F}\{E[s_y s_y^*]\} \quad (7)$$

$$\begin{aligned}\Phi_{ee}=E[\Delta^2]\{(j\omega_x)(-j\omega_x)\cdot\Phi_{ss}(\omega_x,\omega_y)\}+\\ \{(j\omega_y)(-j\omega_y)\cdot\Phi_{ss}(\omega_x,\omega_y)\}.\end{aligned} \quad (8)$$

Simplifying eq.(8) yields the following prediction error psd.

$$\Phi_{ee}(\omega_x,\omega_y)=E[\Delta^2](\omega_x^2+\omega_y^2)\Phi_{ss}(\omega_x,\omega_y). \quad (9)$$

#### *3.1.1. Validation of the time domain problem formulation in frequency domain*

Now let's calculate the prediction error variance directly in frequency domain. Taking the Fourier transform of the expectation of main error equation yields

$$\Phi_{ee}=\Phi_{ss}-2\Re\{\Phi_{bs}\}+\Phi_{bb} \quad (10)$$

where cross power spectral density has the form:

$$\begin{aligned}\Phi_{sb}=\mathcal{F}\{E[s(x,y)s(x-\Delta_x,y-\Delta_y)^*]\}=\\ \Phi_{ss}\cdot E[e^{-j\omega_x\Delta_x-j\omega_y\Delta_y}]\end{aligned} \quad (11)$$

In [4], translational errors are modelled with Gaussian distribution $p(\Delta x,\Delta y)=\frac{1}{2\pi\sigma^2}e^{-\frac{\Delta x^2+\Delta y^2}{2\sigma^2}}$, therefore eq.(11) becomes

$$\Phi_{sb}=\Phi_{ss}e^{-(\omega_x^2+\omega_y^2)\sigma^2/2} \quad (12)$$

Similarly for $\Phi_{bb}$ the following expression is found.

$$\begin{aligned}\Phi_{bb}&=\mathcal{F}\{E[s(x-\Delta_x,y-\Delta_y)s(x-\Delta_x,y-\Delta_y)^*]\}\\ &=\Phi_{ss}E[e^{-j\omega_x\Delta_x-j\omega_y\Delta_y}e^{j\omega_x\Delta_x+j\omega_y\Delta_y}]=\Phi_{ss}\end{aligned} \quad (13)$$

Substituting eq.(13) and eq.(12) in eq.(10) yields:

$$\Phi_{ee}=\Phi_{ss}\cdot 2(1-e^{-(\omega_x^2+\omega_y^2)\sigma^2/2}) \quad (14)$$

Expansion of the exponential in eq.(14) by Maclaurin series with $u=-(\omega_x^2+\omega_y^2)\sigma^2/2$ is $e^u=(1+u+u^2/2!+\ldots)$, and first term approximation similar to eq.(4) gives

$$\Phi_{ee}\approx\Phi_{ss}\cdot 2(-u))=\sigma^2\cdot(\omega_x^2+\omega_y^2)\Phi_{ss} \quad (15)$$

where we reach the same expression as in eq.(9) with $\sigma^2$= $E[\Delta^2]$.

Note that although direct frequency domain derivation gives a neater solution for translational motion errors; it is not straightforward to obtain an exact expression for probabilistic scale and rotational errors. Basically, shift (i.e. translation) in time domain will introduce a phase change in frequency domain, which is separable from the video signal psd. On the other hand, rotational and scale errors affect the frequency components of the video signal psd nonlinearly and because of probabilistic error parameters, it is harder to obtain an analytic expression for prediction error variance directly in frequency domain. Therefore, in the next two subsections prediction error expressions due to scale and rotational errors will be derived starting from time domain.

### 3.2. Scale Error

Let $\Delta a$ be the scale error (equal in both $x$ and $y$ directions). Assuming that the error is due to only scale parameter, then motion compensated block $b(x,y)$ can be written as

$$b(x,y)=s(x-\Delta ax,y-\Delta ay). \quad (16)$$

Taylor series expansion of eq.(16) with the first two terms approximation yields

$$\begin{aligned}s(x-\Delta ax,y-\Delta ay)=s(x,y)+\{-\Delta axs_x-\Delta ays_y\}\\ +\frac{1}{2!}\{\Delta a^2x^2s_{xx}+\Delta a^2y^2s_{yy}+2\Delta a^2xys_{xy}\}\end{aligned} \quad (17)$$

Let $M=xs_x+ys_y$ and $N=x^2s_{xx}+y^2s_{yy}+2xys_{xy}$, where $s_{xx}$, $s_{yy}$ and $s_{xy}$ represent the quantities $(\partial/\partial x)^2 s$,

$(\partial/\partial y)^2 s$, $(\partial^2 s/\partial x \partial y)$ respectively. Then, the prediction error can be written by using the first two terms of approximation as $e(x, y) = \Delta a M - \frac{1}{2}\Delta a^2 N$, which yields the following error psd

$$\Phi_{ee} = E[\Delta a^2]\mathcal{F}\{E[MM^*]\} - E[\Delta a^3]\mathcal{F}\{E[MN^*/2+ NM^*/2]\} + E[\Delta a^4]\mathcal{F}\{E[NN^*]\} \quad (18)$$

If we consider a very good global motion estimator, then the scaling errors are only due to rounding. Therefore, we can assume that the error is uniformly distributed between $-\beta$ and $\beta$ with zero mean, where $\beta$ represents the accuracy level of scale parameter. The higher order statistics of the uniform distribution is $E[\Delta a^2] = \beta^2/3$, $E[\Delta a^3] = 0$, $E[\Delta a^4] = \beta^4/5$ etc. Note that all the odd moments of the uniform distribution will be zero due to symmetry of the uniform distribution around zero, and the higher moments decrease since the inaccuracy level $\beta$ is small. Therefore ignoring the latter terms in the approximation is reasonable in obtaining the following error psd expression:

$$\Phi_{ee} = E[\Delta a^2]\mathcal{F}\{E[MM^*]\} + \frac{E[\Delta a^2]}{4}\mathcal{F}\{E[NN^*]\} \quad (19)$$

with $\mathcal{F}\{E[MM^*]\} = \left|\frac{\partial \omega_x S(\omega_x,\omega_y)}{\partial \omega_x} + \frac{\partial \omega_y S(\omega_x,\omega_y)}{\partial \omega_y}\right|^2$, and $\mathcal{F}\{E[NN^*]\} = \left|\frac{\partial^2 \omega_x^2 S}{\partial \omega_x^2} + \frac{\partial^2 \omega_y^2 S}{\partial \omega_y^2} + 2\frac{\partial^2 \omega_x \omega_y S}{\partial \omega_y \partial \omega_x}\right|^2$.
$S(\omega_x, \omega_y)$ is the Fourier transform of the video signal $s(x, y)$.

### 3.3. Rotational Error

Let $\Delta\theta$ be the error in the rotation parameter. If the motion error is due to only rotation, then $a_{11}$ and $a_{12}$ parameters in eq.(1) will be $\cos\theta$, and $-\sin\theta$ respectively. Since $\Delta\theta$ refers to errors and therefore very small; $\cos\Delta\theta$ and $\sin\Delta\theta$ can be approximated by 1 and $\Delta\theta$ respectively. Then motion compensated block $b(x, y)$ can be written as

$$b(x, y) = s(x + \Delta\theta y, y - \Delta\theta x). \quad (20)$$

Taylor expansion of eq.(20) with first two terms approximation yields

$$s(x + \Delta\theta y, y - \Delta\theta x) = s(x, y) + (\Delta\theta y s_x - \Delta\theta x s_y) + \frac{1}{2!}(\Delta\theta^2 y^2 s_{xx} - 2\Delta^2\theta x y s_{xy} + \Delta\theta^2 x^2 s_{yy}) \quad (21)$$

Let $K = x s_y - y s_x$ and $L = (y^2 s_{xx} - 2xy s_{xy} + x^2 s_{yy})$, then the prediction error is $e(x, y) = \Delta\theta K - \frac{1}{2}\Delta\theta^2 L$ and the error psd is given as

$$\Phi_{ee} = E[\Delta\theta^2]\mathcal{F}\{E[KK^*]\} + \frac{E[\Delta\theta^2]}{4}\mathcal{F}\{E[LL^*]\} \quad (22)$$

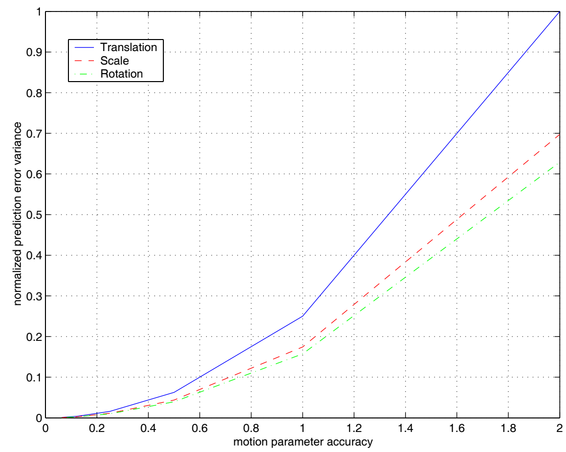


**Fig. 1**. Normalized prediction error variance as a function of global motion parameter inaccuracies.

with $\mathcal{F}\{E[KK^*]\} = \left|\frac{\partial \omega_y S(\omega_x,\omega_y)}{\partial \omega_x} - \frac{\partial \omega_x S(\omega_x,\omega_y)}{\partial \omega_y}\right|^2$, and $\mathcal{F}\{E[LL^*]\} = \left|\frac{\partial^2 \omega_x^2 S}{\partial \omega_y^2} + \frac{\partial^2 \omega_y^2 S}{\partial \omega_x^2} - 2\frac{\partial^2 \omega_x \omega_y S}{\partial \omega_y \partial \omega_x}\right|^2$.

## 4. NUMERICAL RESULTS

In this section, the relation between prediction error variance and the three error (rotational, scale and translational) parameters are evaluated numerically. As in [5], we assume that the video signal $s(x, y)$ has a power spectrum (eq. (23)) that corresponds to a separable autocorrelation function with $R(x, y) = e^{-\alpha(|x|+|y|)}$:

$$\Phi_{ss}(\omega_x, \omega_y) = \frac{4\alpha^2}{(\alpha^2 + \omega_x^2)\cdot(\alpha^2 + \omega_y^2)} \quad (23)$$

$S(\omega_x, \omega_y)$ that is used in eqs.(19, 22) for derivation of $\Phi_{ee}$ is calculated similar to [6]. Fig. 1 depicts the dependency of normalized prediction error variance to motion parameter inaccuracies for $\alpha = 0.9$. The motion errors are uniformly distributed between $[-\beta, \beta]$. The horizontal axis represents $\beta$, where $\beta$ is 0.5 for integer-pel accuracy, $\beta$ is 0.25 for half-pel accuracy and so on. If the gain in dB is defined as $10 log_{10}\sigma^2$, then Fig. 1 suggests that for all three parameters, doubling the accuracy introduces a ≈6 dB gain (or equivalently reduce bit rate by 1 bit/sample). Sub-pixel accuracy becomes more important in translational errors, since in a typical video sequence, the dynamic range of a displacement is bigger than the range of the scaling factor. The ranges in which the global motion parameters fall depends on the characteristics of a video sequence. For example, in a video sequence with moderate motion, the scale change can be in the range [0.9 1.1], where the size of displacement can be much larger. For this reason Fig. 1 does not directly

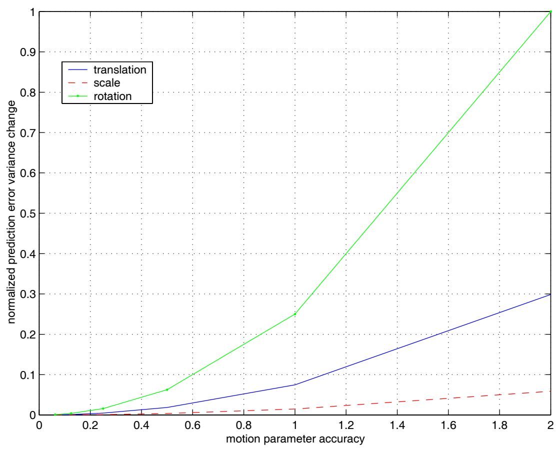


**Fig. 2**. The effect of video signal model

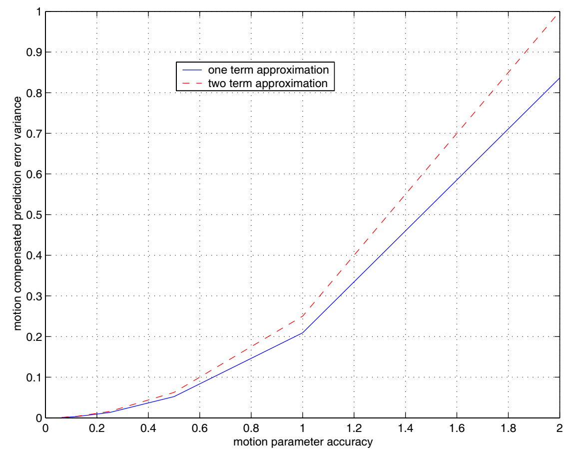


**Fig. 3**. The effect of approximation

suggest that translational motion accuracy is more important than the rotation and scale parameters in increasing the prediction gain for the whole class of video signals.

### 4.1. Effect of signal model assumption

Besides the accuracy level, the video signal characteristics also affect the prediction gain. To understand the effect of video signal, we have analyzed how the prediction error variance due to individual motion parameters change as $\alpha$ in eq.(23) change from 0.22 to 0.55. (examples of $\alpha$ and corresponding images can be found in [5].) Fig. 2 illustrates that the prediction gain that is obtained by changing scale accuracy is less susceptible to video signal model. While prediction gain that can be obtained by changing the rotation accuracy is sensitive and can vary more for different video sequences. This brings us to the point that individual accuracy for each of these motion parameter should be optimized based on the characteristics of the video frame.

### 4.2. Effect of approximation

The effect of approximation in derivation of the prediction error variance is analyzed. Fig. 3 shows how the prediction error variance changes by approximating the psd of error by the first term and by using the first two terms of the taylor's series expansion (see eq.(17)) for scale parameter. As the accuracy level increases, approximation errors also increase. Since, typical range of scale is far below $\beta$=0.5 accuracy, therefore, using the first few terms is reasonable in approximating the exact psd error expression.

## 5. CONCLUSION

In this paper, the effect of different global motion parameter accuracies on the prediction efficiency is studied. It is found that prediction gain is almost linearly proportional to the error variance of the motion parameters, and increasing the accuracy of the motion parameters can reduce the bit rate theoretically up to 1 bit/sample. The effect of these parameters in the efficiency of the video coding depends on the characteristics of the video signal. Future work will be in optimizing the accuracy of global motion parameters for video coding applications.

## 6. REFERENCES


[1] R. Buschmann, "Efficiency of displacement estimation techniques", *Signal Process.: Image Communication*, vol. 10, pp.43-61, 1997.

[2] J. Ribas-Corbera, D.L. Nenhoff, "Optimizing motion-vector accuracy in block-based video coding," *IEEE Trans. on Circuits and Systems for Video Technology*, vol.11, no.4, pp.497-511, April, 2001.

[3] B. Girod,"The efficiency of Motion-Compensating Prediction for Hybrid coding of Video Sequences", *IEEE Journal on Selected Areas in Communications*, vol. sac-5, no. 7, pp. 1140-1154, August, 1987.

[4] B. Girod, "Efficiency analysis of multihypothesis motion-compensated prediction for video coding," *IEEE Trans. on Image Processing*, vol. 9, no. 2, pp. 173-183, 2000.

[5] A. Habibi, A. Wintz, "Image coding by linear transformations and block quantization", *IEEE Trans. on Communication Technology*, vol. 19, pp. 50-62, Feb. 1971.

[6] A. Bruckstein, M. Elad, R. Kimmel,"Down-scaling for better transform compression", *IEEE Trans. on Image Processing*, vol. 12, no. 9, pp. 1132-1144, Sep. 2003.