Rate-Distortion Theory for Simplified Affine Motion Compensation Used in Video Coding

Holger Meuel
Institut für Informationsverarbeitung
Leibniz Universität Hannover
Hannover, Germany
meuel@tnt.uni-hannover.de

Stephan Ferenz
Institut für Informationsverarbeitung
Leibniz Universität Hannover
Hannover, Germany
ferenz@tnt.uni-hannover.de

Yiqun Liu
Institut für Informationsverarbeitung
Leibniz Universität Hannover
Hannover, Germany
liu@tnt.uni-hannover.de

Jörn Ostermann
Institut für Informationsverarbeitung
Leibniz Universität Hannover
Hannover, Germany
office@tnt.uni-hannover.de

***Abstract*—In this work, we derive the rate-distortion function for video coding using the simplified affine, 4-parameter motion compensation model as it is used in the Joint Exploration Model (JEM) by the Joint Video Exploration Team (JVET) on Future Video coding. We model the displacement estimation error during motion estimation and obtain the bit rate by applying the rate-distortion theory. We assume that the displacement estimation error is caused by perturbed parameters of the simplified affine model. These transformation parameters are assumed statistically independent, with each of them having a zero-mean Gaussian distributed estimation error. The joint probability density function (p.d.f.) of the displacement estimation errors is derived and related to the prediction error. We calculate the bit rate as a function of the accuracy of the parameter estimation for the simplified affine motion model. Finally, we compare our results with a translational motion model as used in video coding standards like HEVC as well as with a full affine motion model with 6 degrees of freedom. For aerial sequences containing distinct affine motion, the minimum required bit rate to encode the prediction error can be significantly reduced from 2.5 bit/sample to 0.02 bit/sample for a reasonable operating point and a block size of 64×64 pel^2.**

***Index Terms*—Affine Transformations, Global Motion Compensation, GMC, Rate-Distortion Theory, Efficiency Analysis, Aerial Video Coding, JVET, JEM**

I. Introduction

Motion compensated prediction (MCP) is one of the key elements in modern hybrid video coding standards like *Advanced Video coding* (AVC) [1] or *High Efficiency Video Coding* (HEVC) [2]. MCP is typically performed block-wise for blocks of different sizes, e. g. of 4×4 up to 16×16 pel^2 for AVC or $64{\times}64$ pel^2 for HEVC. The minimum bit rate of the prediction error of motion compensated prediction as a function of the variance of the displacement estimation error was theoretically derived by Girod [3].

Nowadays, new scenarios with distinct global motion—like videos captured from *Unmanned Aerial Vehicles (UAV)*/*Micro Aerial Vehicles (MAV)* like multicopters—emerge and are also considered in recent test sets [4]–[6]. These video sequences contain higher-order global motion, which cannot accurately be described by a purely translational motion model. To cope with such motions better, the *ITU-T/ISO/IEC Joint Video Exploration Team (JEVT)* (on Future Video Coding) incorporated a simplified 4-parameter affine motion model into their reference software *Joint Exploration Model (JEM)* [7], [8]. Affine (as well as homographic) global motion compensation (GMC) is also contained in the video codec AV1 [9]. Early JVET studies based on the initial JEM software (ver. 1.0) on the common test set [10] (containing no sequences with distinct non-translational motion) show coding efficiency gains of up to 1.35 % (JEM 1.0, configuration *Low Delay P (LDP) main 10*) [11], [12]. Larger gains can be expected for sequences containing more higher-order motions like rotation or zoom [13], [14].

Recently, the rate-distortion optimized efficiency of affine global motion compensation has been analyzed in [15]. However, there a full affine model with 6 degrees of freedom was assumed. Since shearing is rare to observe in real data, a simplified model with only 4 degrees of freedom (rotation, scaling, translation) is sufficient to cover most affine motion contained in a scene. Consequently, a simplified model is employed in JEM. For such a model, the assumption of independent estimated affine transformation parameters cannot be met. The effect of global motion parameter inaccuracies employing such a simplified model has been investigated in [16]. In their work, Dane and Nguyen introduced probabilistic rotational, scale and translational errors and derived that by doubling the accuracy of the motion parameter estimates, a theoretical gain up to 6 dB can be obtained in prediction error variance (which corresponds to 1 bit/sample). However, they did not relate their results to other motion models (e. g. purely translational, full affine), albeit the bit rate for encoding the prediction error highly depends on the used motion model. Moreover, the bit rate of the prediction error was not considered in their work at all.

In this work, we present an efficiency analysis including rate-distortion optimization of the simplified affine motion model. We analytically derive the power spectral density of the prediction error after motion compensation as a function of the (simplified) affine transform parameter accuracy in Section II, especially considering dependencies between the parameters of the simplified affine model. Simulations are presented in Section III and Section IV concludes the paper.

II. Efficiency Analysis of the Simplified Affine Model

An efficiency analysis of a full affine motion model has recently been presented in [15]. In contrast to that we assume a simplified affine motion model as used in JEM and thus have to consider dependencies between the parameters of the simplified affine model. First, we model the joint probability density function (p.d.f.) of the displacement estimation error in Subsection II-A before we calculate the bit-rate of the prediction error according to the rate-distortion theory [17] in Subsection II-B.

A. Derivation of the probability density function

We assume a simplified affine model with 4 parameters, as proposed by [13] and used in the JEM software. With the rotation angle θ, the scaling factor s in both horizontal and vertical direction, and the translational parameters c and f, in [13] the relationship between the coordinates x' and y' before and x and y after the transformation is described as

$$\begin{cases} x = s\cos\theta\cdot x' + s\sin\theta\cdot y' + c \ ; \\ y = -s\sin\theta\cdot x' + s\cos\theta\cdot y' + f \ . \end{cases} \tag{1}$$

Replacing $(s\cos\theta)$ and $(s\sin\theta)$ with $(1{+}a)$ and b, respectively, (1) can be rewritten as

$$\begin{cases} x = (a+1)\cdot x' + b\cdot y' + c \ ; \\ y = -b\cdot x' + (a+1)\cdot y' + f \ . \end{cases} \tag{2}$$

We assume that each parameter a, b, c, f is perturbed by an independent error term e_i, with $i = \{a, b, c, f\}$ caused by inaccurate parameter estimation. The perturbed coordinates $\hat{x}$, $\hat{y}$ lead to estimation errors in horizontal and vertical direction of Δx and Δy (in pel)

$$\begin{cases} \Delta x = \hat{x} - x = \quad e_a\cdot x' + e_b\cdot y' + e_c \ ; \\ \Delta y = \hat{y} - y = -e_b\cdot x' + e_a\cdot y' + e_f \ . \end{cases} \tag{3}$$

Assuming each error term e_i to be zero-mean Gaussian distributed leads to the probability density functions (p.d.f.s)

$$p(e_i) = \frac{1}{\sqrt{2\pi\sigma_{e_i}^2}}\cdot\exp\left(-\frac{e_i^2}{2\sigma_{e_i}^2}\right) \tag{4}$$

We assume a Gaussian distribution as the worst-case scenario since it has the maximal entropy of all distributions with the same variance. For statistically independent variables we get a joint p.d.f. $p_{E_a,E_b,E_c,E_f}(e_a, e_b, e_c, e_f)$ for the random variables E_a, E_b, E_c, E_f generating the observations e_a, e_b, e_c, e_f:

$$p_{E_a,E_b,E_c,E_f}(e_a, e_b, e_c, e_f) = p(e_a)\cdot p(e_b)\cdot p(e_c)\cdot p(e_f) \ . \tag{5}$$

To convert the p.d.f. $p_{E_a,E_b,E_c,E_f}(e_a, e_b, e_c, e_f)$ to the desired p.d.f. $p_{\Delta X,\Delta Y}(\Delta x, \Delta y)$ of the displacement estimation errors (in pel) Δx, Δy caused by affine parameter inaccuracies, we use the transformation theorem for p.d.f.s [18]

$$p_{\mathscr{Y}_1,\ldots,\mathscr{Y}_M}(\mathscr{y}_1,\ldots,\mathscr{y}_M) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} p_{\mathscr{X}_1,\ldots,\mathscr{X}_N}(\xi_1,\ldots,\xi_N) \cdot\prod_{m=1}^{M}\delta(\mathscr{y}_m - g_m(\xi_1,\ldots,\xi_N))\mathrm{d}\xi_1\ldots\mathrm{d}\xi_N \ , \tag{6}$$

with $\delta(\cdot)$ denoting the Dirac delta function, $g_1,\ldots,g_M$ being functions $\mathscr{y}_1 = g_1(\mathscr{x}_1,\ldots,\mathscr{x}_N)$, $\ldots$, $\mathscr{y}_M = g_M(\mathscr{x}_1,\ldots,\mathscr{x}_N)$ and $p_{\mathscr{Y}_1,\ldots,\mathscr{Y}_M}(\mathscr{y}_1,\ldots,\mathscr{y}_M)$ being the joint p.d.f. With (3) this yields

$$\begin{aligned} p_{\Delta X,\Delta Y}(\Delta x, \Delta y) = &\int_{\mathbb{R}^4} p_{E_a,\ldots,E_f}(e_a,\ldots,e_f) \\ &\cdot\ \delta(\,\Delta x - (e_a x' + e_b y' + e_c)\,) \\ &\cdot\ \delta(\,\Delta y - (-e_b x' + e_a y' + e_f)\,)\ \mathrm{d}e_a\mathrm{d}e_b\mathrm{d}e_c\mathrm{d}e_f \ , \end{aligned} \tag{7}$$

with a dependency on the location coordinates x', y' in the source frame. Using the properties of the delta function gives

$$\begin{aligned} &p_{\Delta X,\Delta Y}(\Delta x, \Delta y) \\ &= \int_{\mathbb{R}^2} p_{E_a,E_b,E_c,E_f}(e_a, e_b, \Delta x - e_a x' - e_b y', \\ &\quad \Delta y + e_b x' - e_a y')\ \mathrm{d}e_a\mathrm{d}e_b \ . \end{aligned} \tag{8}$$

From (8) with (4) we get

$$\begin{aligned} p_{\Delta X,\Delta Y}(\Delta x, \Delta y) = &\frac{1}{(2\pi)^2\sigma_{e_a}\sigma_{e_b}\sigma_{e_c}\sigma_{e_f}} \\ &\cdot\int_{\mathbb{R}^2}\exp\left(-\frac{e_a^2}{2\sigma_{e_a}^2} - \frac{e_b^2}{2\sigma_{e_b}^2} - \frac{(\Delta x - e_a x' - e_b y')^2}{2\sigma_{e_c}^2}\right. \\ &\left. - \frac{(\Delta y + e_b x' - e_a y')^2}{2\sigma_{e_f}^2}\right)\mathrm{d}e_a\mathrm{d}e_b \ . \end{aligned} \tag{9}$$

After the two integrations we obtain

$$p_{\Delta X,\Delta Y}(\Delta x, \Delta y) = \frac{1}{2\pi\sqrt{N}}\cdot\exp\left(\frac{M}{2N}\right) \tag{10}$$

$$\begin{aligned} \text{with } N = &\left(\left(x'^2 + y'^2\right)^2\sigma_{e_b}^2 + y'^2\sigma_{e_c}^2 + x'^2\sigma_{e_f}^2\right)\sigma_{e_a}^2 \\ &+ \left(x'^2\sigma_{e_c}^2 + y'^2\sigma_{e_f}^2\right)\sigma_{e_b}^2 + \sigma_{e_c}^2\sigma_{e_f}^2 \ , \end{aligned} \tag{11}$$

$$\begin{aligned} \text{and } M = &-(x'\Delta y - y'\Delta x)^2\,\sigma_{e_a}^2 - (x'\Delta x + y'\Delta y)^2\,\sigma_{e_b}^2 \\ &- \Delta x^2\sigma_{e_f}^2 - \Delta y^2\sigma_{e_c}^2 \ . \end{aligned} \tag{12}$$

Transforming (10) into the form of a common bivariate zero-mean normal distribution with ρ being the correlation coefficient between ΔX and ΔY leads to the desired final p.d.f. of the displacement estimation error

$$\begin{aligned} p_{\Delta X,\Delta Y}(\Delta x, \Delta y) = &\frac{1}{2\pi\sigma_{\Delta X}\sigma_{\Delta Y}\sqrt{1-\rho^2}} \\ &\cdot\exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{\Delta x^2}{\sigma_{\Delta X}^2} + \frac{\Delta y^2}{\sigma_{\Delta Y}^2} - \frac{2\rho\cdot\Delta x\cdot\Delta y}{\sigma_{\Delta X}\cdot\sigma_{\Delta Y}}\right]\right) \end{aligned} \tag{13}$$

$$\text{with } \sigma_{\Delta X}^2 = N\cdot\left(\left(\sigma_{e_a}^2 y'^2 + \sigma_{e_b}^2 x'^2 + \sigma_{e_f}^2\right)\cdot\left(1-\rho^2\right)\right)^{-1}, \tag{14}$$

$$\sigma_{\Delta Y}^2 = N\cdot\left(\left(\sigma_{e_a}^2 x'^2 + \sigma_{e_b}^2 y'^2 + \sigma_{e_c}^2\right)\cdot\left(1-\rho^2\right)\right)^{-1}, \tag{15}$$

$$\rho = \frac{\left(\sigma_{e_a}^2 x'y' - \sigma_{e_b}^2 x'y'\right)}{\sqrt{\sigma_{e_a}^2 y'^2 + \sigma_{e_b}^2 x'^2 + \sigma_{e_f}^2}\sqrt{\sigma_{e_a}^2 x'^2 + \sigma_{e_b}^2 y'^2 + \sigma_{e_c}^2}}. \tag{16}$$

As can be seen, the variances $\sigma_{\Delta X}^2$ and $\sigma_{\Delta Y}^2$ depend on the locations x', y'. Moreover, the variances of Δx and Δy both depend on the variances of *all* estimated parameters and thus the underlying random processes ΔX and ΔY are dependent.

For equal variances $\sigma^2_{e_a} = \sigma^2_{e_b}$, ρ becomes zero.

B. Rate-distortion analysis

To derive the bit rate for coding the prediction error in motion compensated video coding, we use the findings from Girod, who related the displacement estimation error $p_{\Delta X,\Delta Y}(\Delta x, \Delta y)$ to the prediction error e_{p} [3]. Applying the rate-distortion theory [17] results in the minimum achievable bit rate for encoding the prediction error. In this subsection we will summarize the derivations from [3].

Given a displacement estimation error $p_{\Delta X,\Delta Y}(\Delta x, \Delta y)$, we obtain the power spectral density of the prediction error

$$S_{ee}(\Lambda) = 2\, S_{ss}(\Lambda)\, [1 - \mathrm{Re}(P(\Lambda)] + \Theta \;, \tag{17}$$

where $S_{ss}(\Lambda)$ denotes the power spectral density of the video signal s, Λ being the two-dimensional (2D) spatial frequency vector $\Lambda := (\omega_x, \omega_y)$, $P(\Lambda)$ being the 2D Fourier transform of the probability density function (p.d.f.) of the displacement estimation error, and Θ being a parameter that generates the function $R(D)$ by taking on all positive real values ([3], equation (28)). $S_{ss}(\omega_x, \omega_y)$ was determined according to O'Neil [19] and Girod [3], where the statistics of the source was assumed to be represented by the autocorrelation function

$$R_{ss}(\Delta x, \Delta y) = E\,[s(x,y)\cdot s(x-\Delta x, y-\Delta y)] := \exp\left(-\alpha\,\sqrt{\Delta x^2 + \Delta y^2}\right). \tag{18}$$

We assume α not to be isotropic and thus replace (18) by $\exp\left(-\sqrt{\alpha_x^2 \Delta x^2 + \alpha_y^2 \Delta y^2}\right)$. The exponential drop rates α_x and α_y in x- and y-direction can be determined as the negative logarithm of the correlations between horizontally and vertically adjacent pixels $\alpha_x = -\ln(\rho_x)$ and $\alpha_y = -\ln(\rho_y)$ [19]. For this, the Pearson correlation coefficients $\rho(X,Y) = \frac{\mathrm{cov(X,Y)}}{\sigma_X \sigma_Y}$ and similarly ρ_Y with the standard deviations σ_X, σ_Y and the covariance cov were determined [20]. The desired power spectral density $S_{ss}(\Lambda)$ to be inserted in (17) is now the Fourier transform of (18).

Finally, we derive the distortion D as well as the corresponding minimum transfer rate $R(D)$ from the rate-distortion function for a given mean-squared error ([3], equations 19–20)

$$D = \frac{1}{4\pi^2} \iint_{\Lambda} \min\,[\Theta, S_{ss}(\Lambda)]\, \mathrm{d}\Lambda \;, \tag{19}$$

$$R(D) = \frac{1}{8\pi^2} \iint_{\substack{\Lambda:(S_{ss}(\Lambda)>\Theta \\ \text{and } S_{ee}(\Lambda)>\Theta)}} \log_2\left[\frac{S_{ee}(\Lambda)}{\Theta}\right] \mathrm{d}\Lambda \;\text{ bit} \;. \tag{20}$$

We would like to emphasize that our $\sigma^2_{\Delta X}$ and $\sigma^2_{\Delta Y}$ are location dependent, since they are functions of the source pixel coordinates x', y'. Consequently, $p_{\Delta X,\Delta Y}(\Delta x, \Delta y)$, $P(\Lambda)$ and $S_{ee}(\Lambda)$ are also location dependent.

Using the idea of generating the rate-distortion function for translative motion like explained by Girod [3] and our results from Section II, we derived the rate-distortion function for the simplified affine motion model as defined in (2).

III. Simulations

In our simulations, we evaluate the minimal bit rate for simplified affine global motion compensated prediction. As we have seen in the last section, the variance of the displacement estimation error $p_{\Delta X,\Delta Y}(\Delta x, \Delta y)$ depends on the location in the image according to (14) and (15). Thus, the resulting minimum achievable bit rate is location dependent. To obtain the total bit rate for encoding one frame, we average the pel-wise bit rates afterwards.

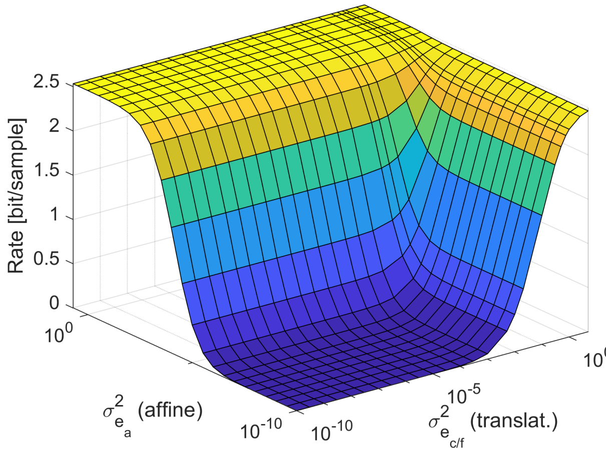
For calculating the power spectral density S_{ss} of the video signal in (17) and the distortion in (19), we used the exponential drop rates $\alpha_x = 0.9744$ and $\alpha_y = 0.9677$ of the autocorrelation function (eq. (18)) as measured in [15].

Evaluation of the rate-distortion theory for a distortion of SNR = 30 dB results in minimum required bit rates for different variances $\sigma^2_{e_i}$ of Gaussian displacement estimation error p.d.f.s of the affine transform parameters as shown in Fig. 1. For the simulations we assumed the affine parameters to be in a fixed ratio ($\sigma^2_{e_b} = 2\sigma^2_{e_a}$) and both translational parameters to be equal ($\sigma^2_{e_c} = \sigma^2_{e_f}$). The minimum bit rates as a function of the affine and translational variances are presented in Fig. 1a. The relationship between $\sigma^2_{e_a}$ and $\sigma^2_{e_b}$ is justified by the fact that small rotation angles ($\Theta \leq 5°$) are more likely to occur. Then, exploiting the small-angle approximation, $s\cos\Theta$ and $s\sin\Theta$ from (1) approximately become s and $s\cdot\Theta$, respectively. Assuming $s=1$, we get $e_a = \hat{a} - a = \cos\hat{\Theta} - \cos\Theta = -2\sin(\frac{\hat{\Theta}+\Theta}{2})\cdot \sin(\frac{\hat{\Theta}-\Theta}{2}) \approx -2\sin(\Theta)\cdot\sin(\frac{1}{2}\Delta\Theta)$ and accordingly $e_b = \hat{b} - b = \sin\hat{\Theta} - \sin\Theta \approx 2\cos(\Theta)\cdot\sin(\frac{1}{2}\Delta\Theta)$ for $\hat{\Theta} \approx \Theta$. Assuming small Θ, we get $e_a \approx -2\Theta k$ and $e_b \approx 2k$, with $k = \sin(\frac{1}{2}\Delta\Theta)$ being constant. Exploiting the definition of the variance $\sigma^2_{e_i} = \int_{-\infty}^{\infty} p(e_i)(e_i - E\{e_i\})^2 \mathrm{d}e_i$ and using e_a and e_b as derived above, for small angles we get $\sigma^2_{e_a} < \sigma^2_{e_b}$.

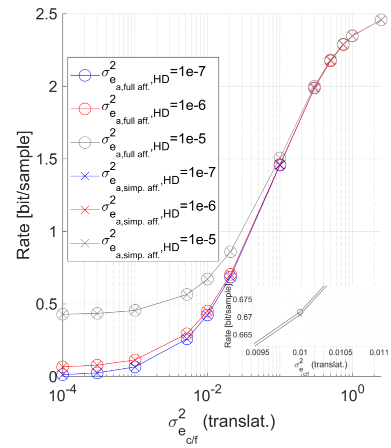
The variances measured for aerial videos from the TAVT data set [6], [21], employing the simplified affine model, confirm the relationship with $\sigma^2_{e_a} = 2.3\mathrm{e}{-7}$, $\sigma^2_{e_b} = 4.6\mathrm{e}{-7}$ on average.

It is noteworthy that the operating point $\sigma^2_{e_a}$, $\sigma^2_{e_{c/f}}$ moves towards significant higher bit rates, if the motion contained in the sequence cannot be represented by the motion model, i. e. if a purely translational motion model is used to estimate a sequence containing distinct (non purely translational) affine motion. For instance, the minimum required bit rate for an accurate simplified affine estimation of $\sigma^2_{e_a} = \sigma^2_{e_{c/f}} = 3\mathrm{e}-7$ amounts 0.019 bit/sample. In contrast to this, for a purely translational motion model, the affine part of the motion contained in the scene cannot be covered at all, leading to high $\sigma^2_{e_a}$, $\sigma^2_{e_b}$ and consequently high bit rates of about 2.5 bit/sample (most right plateau in Fig. 1a).

In Fig. 1b the bit rates are compared for a full affine model (6 degrees of freedom) (circles) as in [15] and a simplified, 4-parameter model (crosses) as analyzed in this paper for $64{\times}64\,\mathrm{pel}^2$ blocks. From the plots we see that the simplified model requires a smaller amount of bits for encoding the prediction error compared to a full affine model for estimations with equal variances as expected. However, the difference is



(a) Block size $64 \times 64\,\mathrm{pel}^2$.



(b) Simplified vs. full affine ($64 \times 64\,\mathrm{pel}^2$) with magnification (bottom right).

Figure 1. Minimum required bit rate versus variances $\sigma^2_{e_i}$, $i = a, b, c, f$ of Gaussian displacement estimation error p.d.f.s for a distortion of SNR = 30 dB assuming $\sigma^2_{e_b} = 2\sigma^2_{e_a}$ and $\sigma^2_{e_c} = \sigma^2_{e_f}$. The surface in (a) shows rates for a block size of $64 \times 64\,\mathrm{pel}^2$) and the transform center in the center of the block, the 2D cuts in (b) represent achievable gains of our simplified affine vs. the full affine motion model from [15] for the same block size.

negligible in terms of bit rate saving. Thus, we recommend to use the simplified affine model for encoding purposes.

IV. Conclusion

In our paper we derive the minimum required bit rate for encoding the prediction error using a simplified affine motion model with 4 degrees of freedom for motion compensated prediction by applying the rate-distortion theory.

We compare the results for the simplified affine motion model with only 4 degrees of freedom with a full affine motion model (6 degrees of freedom) and analyzed that the coding efficiency can be slightly increased by use of the simplified motion model. For encoding the prediction error the minimum required bit rate can be highly reduced from 2.5 bit/sample to about 0.2 bit/sample for sequences containing distinct affine (non-translational) motion, e. g. in videos captured from a UAV, and a reasonable operating point. This, as a consequence, shows, that the usage of the simplified affine model in JEM is justified.

References

[1] AVC, *Rec. ITU-T H.264 & ISO/IEC 14496-10 MPEG-4 Pt.10: Adv. Video Cod. (AVC)-3rd Ed.* Geneva, Switzerland: ISO/IEC&ITU-T, Jul. 2004.

[2] HEVC, "ITU-T Recommendation H.265/ ISO/IEC JTC 1/SC 29 23008-2:2015-05-01 MPEG-H Part 2/: High Efficiency Video Coding (HEVC), 2nd Edition," Apr. 2015.

[3] B. Girod, "The efficiency of motion-compensating prediction for hybrid coding of video sequences," *Selected Areas in Communications, IEEE Journal on*, vol. 5, no. 7, pp. 1140 – 1154, Aug 1987.

[4] X. Zheng, Z. Cao, and F. Wolf, "Aerial photography sequences for video coding standard development," *Joint Video Exploration Team (JVET) of ITU-T VCEG and ISO/IEC MPEG, Chengdu, China, Doc. JVET-D0060*, Oct 2016. [Online]. Available: http://phenix.it-sudparis.eu/jvet/

[5] X. Zheng, W. Li, Z. Cao, W. Su, C. Zhao, Y. Li, Z. Lorenz, H. Wu, Z. Du, and D. A. Hoang, "New aerial photography sequences for video coding standard development," *Joint Video Exploration Team (JVET) of ITU-T VCEG and ISO/IEC MPEG, Hobart, Australia, Doc. JVET-F0062*, Mar 2017. [Online]. Available: http://phenix.it-sudparis.eu/jvet/

[6] Inst. f. Informationsverarb. (TNT), Leibniz Universität Hannover. (2010–2014) TNT Aerial Video Testset (TAVT). [Online]. Available: https://www.tnt.uni-hannover.de/project/TNT_Aerial_Video_Testset/

[7] JVET, "Joint Exploration Model (JEM) of the Joint Video Exploration Team (on Future Video coding) of ITU-T VCEG and ISO/IEC MPEG (JVET)." [Online]. Available: https://jvet.hhi.fraunhofer.de/

[8] J. Chen, E. Alshina, G.-J. Sullivan, J.-R. Ohm, and J. Boyce, "Algorithm Description of Joint Exploration Test Model (JEM) 1," *Joint Video Exploration Team (JVET) of ITU-T VCEG and ISO/IEC MPEG, Geneva, CH, Doc. JVET-A1001*, Oct 2015. [Online]. Available: http://phenix.it-sudparis.eu/jvet/

[9] AOMedia Video 1 (AV1). (2017, Jul.) AOM – AV1: How does it work? [Online]. Available: https://parisvideotech.com/wp-content/uploads/2017/07/AOM-AV1-Video-Tech-meet-up.pdf

[10] K. Suehring and X. Li, "JVET common test conditions and software reference configurations," *Joint Video Exploration Team (JVET) of ITU-T VCEG and ISO/IEC MPEG, 2nd Meeting: San Diego, US, from Feb. 22–26, 2016, Document: JVET-B1010*, Feb 2016. [Online]. Available: http://phenix.it-sudparis.eu/jvet/

[11] H. Zhang, H. Chen, X. Ma, and H. Yang, "Performance analysis of affine inter prediction in JEM1.0," *Joint Video Exploration Team (JVET) of ITU-T VCEG and ISO/IEC MPEG, San Diego, USA, Doc. JVET-B0037*, Feb 2016. [Online]. Available: http://phenix.it-sudparis.eu/jvet/

[12] E. Alshina, A. Alshin, K. Choi, and M. Park, "Performance of JEM 1 tools analysis," *Joint Video Exploration Team (JVET) of ITU-T VCEG and ISO/IEC MPEG, San Diego, USA, Doc. JVET-B0022*, Feb 2016. [Online]. Available: http://phenix.it-sudparis.eu/jvet/

[13] L. Li, H. Li, D. Liu, Z. Li, H. Yang, S. Lin, H. Chen, and F. Wu, "An Efficient Four-Parameter Affine Motion Model for Video Coding," *IEEE Transact. on Circ. and Syst. f. Vid. Techn.*, vol. PP, no. 99, pp. 1–1, 2017.

[14] L. Li, H. Li, Z. Lv, and H. Yang, "An affine motion compensation framework for high efficiency video coding," in *IEEE Internat. Symp. on Circuits and Systems (ISCAS)*, May 2015, pp. 525–528.

[15] H. Meuel, S. Ferenz, Y. Liu, and J. Ostermann, "Rate-Distortion Theory for Affine Global Motion Compensation in Video Coding," in *25rd IEEE Int. Conf. on Image Processing (ICIP)*, Oct. 2018.

[16] G. Dane and T. Q. Nguyen, "The Effect of Global Motion Parameter Accuracies on the Efficiency of Video Coding," in *Int. Conf. on Image Proc. (ICIP)*, vol. 5, Oct 2004, pp. 3359–3362 Vol. 5.

[17] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, ser. Prentice-Hall Elect. Eng. Ser. Prentice-Hall, 1971.

[18] P. Z. Peebles, Jr., *Probability, random variables and random signal principles*. New York: McGraw-Hill Education, 1993, english.

[19] J. O'Neal and T. Natarajan, "Coding Isotropic Images," *IEEE Transactions on Information Theory*, vol. 23, no. 6, pp. 697–707, Nov. 1977.

[20] Royal Society (Great Britain), *Proceedings of the Royal Society of London*. Taylor & Francis, 1895, no. Bd. 58. [Online]. Available: https://books.google.de/books?id=60aL0zlT-90C

[21] H. Meuel, M. Munderloh, M. Reso, and J. Ostermann, "Mesh-based Piecewise Planar Motion Compensation and Optical Flow Clustering for ROI Coding," in *APSIPA Transact. on Sig. and Inform. Proc.*, vol. 4, 2015. [Online]. Available: http://journals.cambridge.org/article_S2048770315000128