ESTIMATING SIDE-INFORMATION FOR WYNER-ZIV VIDEO CODING USING RESOLUTION-PROGRESSIVE DECODING AND EXTENSIVE MOTION EXPLORATION

Wei Liu, Lina Dong and Wenjun Zeng

Computer Science Department, University of Missouri, Columbia, Missouri 65211, USA

ABSTRACT

In Wyner-Ziv video coding (WZVC), the quality of side information (SI) has a critical impact on the coding efficiency. Most existing WZVC schemes generate SI using decoder-side motion estimation (ME), where the current frame is unavailable and the ME accuracy is greatly impaired. In this paper, without introducing additional bitrate overhead, we incorporate fractional-pel motion search, reduced block sizes and multiple hypothesis prediction into our previously proposed decoder-side multi-resolution motion refinement framework [1], where the current frame is progressively decoded, and based on which the decoder iteratively refines the motion to improve its accuracy. Theoretical analysis shows significant gain of the combination of these advanced techniques. A practical SI estimator is implemented and provides prediction performance comparable to H.264/AVC.

Index Terms—Wyner-Ziv video coding, motion estimation, multi-resolution processing, rate-distortion analysis.

1. INTRODUCTION

Wyner-Ziv video coding (WZVC) is gaining increasing attention because of its important applications in video surveillance, low-cost cameras and wireless sensor networks, etc. In WZVC, frames are encoded as syndromes, and the syndromes are decoded based on some side information (SI) available at the decoder. The quality of the SI is critical to the rate-distortion (R-D) performance of WZVC. The SI is usually generated from motion compensated prediction (MCP) based on previously decoded frames, thus accurate motion estimation (ME) is crucial to better SI generation. In most application scenarios of WZVC, encoder-side ME is prohibited by its complexity, thus ME has to be performed at the decoder side. One significant difference between decoder-side ME and encoder-side ME is that the decoder does not have access to the current frame. This hurts the accuracy of the estimated motion vectors (MV), and consequently, more bits (syndromes) are needed to reconstruct the current frame. This rate loss is referred to as the video coding loss in [2], which has been the bottleneck in improving the coding efficiency in WZVC.

If the decoder does not have any access to the current frame, temporal domain motion extrapolation is usually employed to derive the motion for the current frame. The accuracy of motion extrapolation is poor. Analytical results in [2] suggest that WZVC with motion extrapolation could fall 6dB or more behind conventional video coding due to inaccurate MCP.

Many schemes have been proposed to improve the decoder-side ME accuracy for WZVC. One common feature of these approaches is to let decoder have partial access to the current frame, which could be the hash code [3] or CRC code [4] of selected blocks, or some partially-decoded form of the current frame [5][6]. In our previous work [1] we propose to decode the current frame progressively in the resolution dimension. Once a low-resolution version of the current frame is reconstructed, ME is performed with respect to previously decoded frame(s) to refine the motion field, and the refined motion information is employed in the Wyner-Ziv decoding of the next resolution level of the frame. This approach is called multi-resolution motion refinement (MRMR). Theoretical analysis shows that, if the same block matching algorithm (BMA) is employed for both MRMR and conventional inter-frame ME (where the estimator has full access to the current frame), MRMR falls only about 1.5 dB behind conventional ME, which is a significant improvement over motion extrapolation.

It should also be noted that unlike encoder-side ME, decoder-side ME does not suffer from the overhead in transmitting the motion information. Hence, a natural question to ask is: can we improve the SI quality of MRMR by providing a more detailed description of the motion field? Conventionally, if a BMA is used for ME, greater details of the motion can be provided by fractional-pel motion search, using smaller block sizes, or using multiple MVs for one block. In this paper, we study the performance of MRMR with these advanced ME techniques integrated. The rest of the paper is organized as follows. Section 2 reviews our previous work on the R-D modeling for MRMR. Section 3 analyzes the potential gain when MRMR is combined with advanced ME techniques. Section 4 presents a practical wavelet-domain SI estimator and the simulation results. Section 5 concludes the paper.

2. REVIEW OF PRIOR WORKS

Girod [7] shows that for a 2-D colored signal s and its MCP residual e, the following relationship holds

$$\Phi_{ee}(\omega) \approx \Phi_{ss}(\omega)\left[1-\exp\left(-\omega^T\omega\sigma^2_{\Delta d}\right)\right] \qquad (1)$$

where Φ_{ss} and Φ_{ee} denote the power spectrum density (PSD) of s and e, respectively, $\omega = (\omega_x, \omega_y)^T$ is the 2D spatial frequency, Δd denotes the error between the estimated MV and the true MV, and $\sigma^2_{\Delta d}$ denotes the variance of "motion displacement". At high rates, the rate saving obtained by MCP over intra-frame coding is [1]

$$\Delta R = R_{MCP} - R_{intra} \approx \frac{1}{8\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\log_2\left[1-\exp\left(-\omega^T\omega\sigma^2_{\Delta d}\right)\right]d\omega \cdot \qquad (2)$$

Eq. (2) suggests that the coding gain of MCP depends exclusively on $\sigma^2_{\Delta d}$, or the accuracy of ME.

[1] In this section and the next, we do not consider the extra bits spent for MVs. However, we should keep in mind that in conventional inter-frame coding, the overhead in transmitting MVs is not trivial.

978-1-4244-2354-5/09/$25.00 ©2009 IEEE

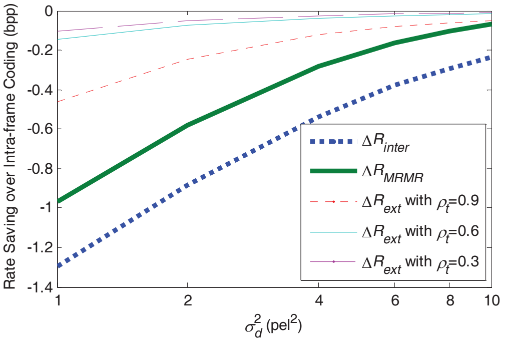

Fig. 1. Comparison of the rate saving performance among inter-frame ME, MRMR and motion extrapolation.

For encoder-side inter-frame ME, BMAs are widely adopted. An excellent model is proposed by Buschmann [10] to estimate the accuracy for BMAs, where a BMA is approximated by a series of data processing applied to the true motion field: estimation filtering, sampling, quantization and reconstruction filtering. If the cost of transmitting the motion information is ignored (i.e., we can use dense sampling to avoid aliasing and use fine fractional-pel motion search to reduce quantization noise), the only operation that contribute to the motion displacement is the estimation filtering, which is caused by using a local neighborhood during the matching. According to our previous result [1], we model the estimating filter as a low-pass filter with the bandwidth π/B, where B is the 1-D block size for matching, and derive the motion accuracy of inter-frame ME as

$$\sigma_{\Delta d}^2 = 2\sqrt{2}\pi^{-2}\ln\rho_s^{-1}B\sigma_d^2 = kB\sigma_d^2 \tag{3}$$

where ρ_s stands the for motion correlation of neighboring pixels. According to [10], we use $\rho_s = 0.983$ for CIF sequences and $\rho_s = (0.983)^2 = 0.966$ for QCIF sequences. The corresponding $k = 0.005$ and 0.01, respectively.

Also in [1], we derive the motion accuracy of MRMR as

$$\sigma_{\Delta d-MRMR}^2(\omega) = \pi kB\sigma_d^2 \Big/ \max\left(\left|\omega_x\right|,\left|\omega_y\right|\right) \cdot \tag{4}$$

Note that in MRMR, if the same block size is used for a down-sampled version, the actual bandwidth of the estimation filter becomes larger. That is why the motion error is a function of ω.

Substitute (3) and (4) into (2), we can get the efficiency of inter-frame ME and MRMR, respectively.

Assume $B = 4$, $k = 0.01$, numerical results are generated and plotted in Fig. 1 for ΔR_{inter} and ΔR_{MRMR}. For comparison purpose, we also plot the rate saving performance of motion extrapolation (ΔR_{ext}) in Fig. 1, using the model constructed in [2], where ρ_t is the temporal motion correlation. We can see that for motion extrapolation, the rate saving performance becomes marginal for sequences with low or medium ρ_t. Even when ρ_t is as high as 0.9, MRMR can still save 0.02 to 0.51 bpp more than motion extrapolation. The gap between ΔR_{MRMR} and ΔR_{inter} is almost a constant around 0.25 bpp. It is well known that for high rate coding, the rate difference at 1 bpp can be translated into 6.02 dB PSNR difference. So we conclude that MRMR falls about 1.5 dB behind inter-frame ME, and outperforms motion extrapolation by up to 5 dB (for medium or low ρ_t).

3. MRMR WITH EXTENSIVE MOTION EXPLORATION

The analysis in the previous section is based on the assumption that both inter-frame ME and MRMR use the same settings in the BMA. However, since MRMR is performed at the decoder side, it has the advantage that the motion information does not have to be transmitted. Without this overhead, it is possible to exploit more advanced ME techniques to extensively explore the dependency between the current frame and the reference frame(s). In the literature, better MCP can be achieved through fine fractional-pel motion search [8], using smaller block sizes or multiple hypothesis prediction [9][11]. In this section, we will analyze the performance of MRMR with these techniques.

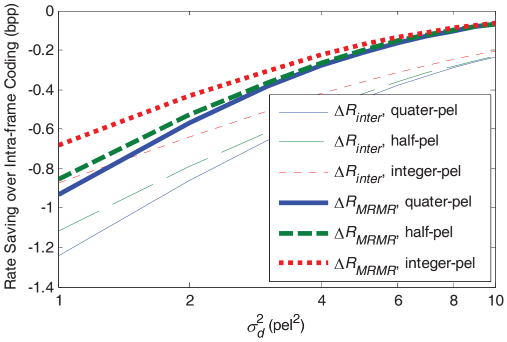

Fig. 2. Rate saving performance of MRMR and inter-frame ME using different pel-accuracy search methods.

3.1 MRMR with fractional-pel motion search

BMAs introduce quantization noise to the true motion field because the reference frame(s) is sampled on a discrete grid. In (3) and (4), the quantization noise is not considered. In this subsection, we model the quantization error in BMAs as an additive white noise with variance $\sigma_{\Delta d-q}^2$, thus (3) and (4) are adjusted as

$$\sigma_{\Delta d-inter}^2 = kB\sigma_d^2 + \sigma_{\Delta d-q}^2 \tag{5}$$

and

$$\sigma_{\Delta d-MRMR}^2(\omega) = \pi kB\sigma_d^2 \Big/ \max\left(\left|\omega_x\right|,\left|\omega_y\right|\right) + \sigma_{\Delta d-q}^2 \tag{6}$$

respectively. According to [10], $\sigma_{\Delta d-q}^2$ is derived by applying uniform scalar quantization with the step size q to a random MV with the p.d.f. $p_d(d)$. Here q represents the pel-accuracy of the BMA (for example, $q = 1$ for integer-pel motion search). We use the same settings as in [10], where $p_d(d)$ is assumed to be a generalized Gaussian distribution with the shape factor being 0.3.

Still assume $B = 4$ and $k = 0.01$, we plot the rate saving curves in Fig. 2 for integer-pel, half-pel, quarter-pel accuracy search for both inter-frame ME and MRMR. It can be seen that the rate saving using fractional-pel accuracy search in MRMR is less significant than in inter-frame ME. This can be explained from (6) that the impact of low-pass filtering noise is more significant in the MRMR case. We also conclude that in MRMR, it is more effective to perform fractional-pel motion search at the high frequency subbands where the low-pass filtering noise is less dominant.

The results in [2] show that improving the pel search accuracy does not help much in motion extrapolation. However, it is usually worthwhile to perform fractional-pel motion search in MRMR. The possible rate saving is up to 0.25 bpp if an integer-pel search is substituted by a quarter-pel search, which can be translated into 1.5 dB in PSNR gain. On the other hand, putting Fig. 1 and Fig. 2 side-by-side we can see, the extra gain is marginal by employing motion search that is finer than quarter-pel.

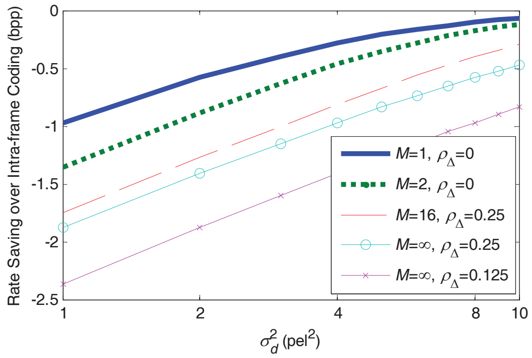

Fig. 3. Rate saving performance of MRMR with MHP.

3.2 MRMR with smaller block sizes

For very accurate ME with small $\sigma^2_{\Delta d}$, using a Taylor series expansion, (2) can be approximated as

$$\Delta R \approx \frac{1}{8\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\log_2\left(\omega^T\omega\sigma^2_{\Delta d}\right)d\omega \cdot \qquad (7)$$

By substituting (4) into (7) we can see that reducing B by half also halves the motion error, which means an extra 0.5 bpp can be saved, or equivalently, the SI quality can be improved by 3 dB. This encourages the use of smaller block sizes. However, block matching with a very small B is an ill-posed problem. People usually impose some smoothness constraint to make sure the derived motion field is physically meaningful. This is equivalent to applying additional inter-block low-pass filtering to the motion field, which somewhat limits the gain of reducing the block size.

3.2 MRMR with multiple-hypothesis prediction

In BMAs, multiple-hypothesis prediction (MHP) is usually employed to improve the prediction performance. One typical example is the bi-directional prediction. In this paper, we will not consider B pictures. In stead, the prediction of the current block is the weighted average of N motion-compensated blocks (each of which is called a hypothesis) from previously decoded frame(s). Using the results in [9], if all the hypotheses are noise-free, the rate saving using MHP is

$$\Delta R_{MHP} = \frac{1}{8\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\log_2\left[1-\mathbf{P}^H E\left(\mathbf{D}\mathbf{D}^H\right)^{-1}\mathbf{P}\right]d\omega \qquad (8)$$

where Δd_1, …, Δd_N denote the MV displacements for the N hypotheses, $\mathbf{D} = [\exp(-\omega^T\Delta d_1), \dots, \exp(-\omega^T\Delta d_N)]^T$, $\mathbf{P} = E(\mathbf{D})$, and the superscript ^{H} stands for the transposed conjugate.

Following the same settings as in [11], we let the motion displacements be jointly Gaussian, each of which has the same variance (denoted as σ^2_Δ), and the correlation between any two displacements are the same (denoted as ρ_Δ). Then (8) can be simplified as (see the Appendix for more details):

$$\Delta R_{MHP} = \frac{1}{8\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\log_2\left[1-\frac{N\exp\left(-\omega^T\omega\sigma^2_\Delta\right)}{1+(N-1)\exp\left(-(1-\rho_\Delta)\omega^T\omega\sigma^2_\Delta\right)}\right]d\omega \cdot \qquad (9)$$

The results in [11] is under the assumption that the N hypotheses are simply averaged, while (9) is the optimum case where the N hypotheses are Wiener filtered.

For very accurate ME, (9) can be approximated as

$$\Delta R_{MHP} = \frac{1}{8\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\log_2\left[\frac{N-1}{N}\left(\frac{1}{N-1}+\rho_\Delta\right)\omega^T\omega\sigma^2_\Delta\right]d\omega \cdot \qquad (10)$$

We can see that for large Ns, reducing ρ_Δ is equally important as reducing σ^2_Δ; if the displacements are mutually independent (ρ_Δ=0), [2] doubling the number of hypotheses is equivalent to reducing σ^2_Δ by half (which means a 3 dB gain in prediction performance); however, while ρ_Δ>0, no significant gain can be obtained by increasing N when N >>1/ρ_Δ.

Eq. (9) is for conventional inter-frame coding, however, it can be easily extended to MRMR by replacing σ^2_Δ with $\sigma^2_\Delta(\omega)$. Let B=4, k=0.01 and assume there is no quantization noise in the ME. We plot the rate-saving curves for MRMR in Fig. 3. Our discussions above can be well confirmed.

4. SIMULATION RESULTS

From the previous analysis, we know that if the same BMA settings are used, there is a 1.5 dB gap between MRMR and inter-frame ME. However, either reducing the block size by half or doubling the number of hypotheses can provide a rate saving by up to 0.5 bpp, making it possible to make up the 1.5 dB gap. Certainly those techniques can also be employed in inter-frame ME, but the corresponding motion overhead is not negligible. For example, the number of MVs will be doubled if the number of hypotheses is doubled, or quadrupled if the block size is halved. State-of-the-art video compression standard H.264/AVC [12] employs quarter-pel search, no smaller than 4x4 blocks and 1 hypothesis (1 MV) for each inter-predictive block. In this section, we will integrate the ME techniques into a practical MRMR framework to reduce the video coding loss, and compare its prediction performance with H.264/AVC.

We use DWT for the multi-resolution decomposition. The reference frame(s) are decomposed using redundant DWT to overcome the shift-variance problem. In the lowest resolution level, the motion field is set to zero. After one resolution level is decoded, motion is refined based on this level and the corresponding level(s) of the reference frame(s). The process iterates until the full resolution frame is decoded. Due to space limit, the reader is referred to [1] for greater details of the system.

We divide the current (low-resolution) frame into 2x2 blocks and perform motion search on quarter-pel accuracy. A smoothness constraint is imposed to avoid a chaotic motion field. Up to 8 hypotheses are searched for each block and their average is used for the prediction (see [11] for details). If no match is found, the prediction is set to zero. For simplicity, only 1 previously reconstructed frame is used as reference.

To measure the prediction performance, we treat the residual samples in a subband as i.i.d., and calculate the rate saving as

$$\Delta R = \frac{1}{2}\log_2\frac{MSE_1}{MSE_0} \qquad (11)$$

where MSE_1 is the mean-square-error (MSE) of the MCP, MSE_0 is the MSE of all-zero prediction (or if intra-frame coding is employed). Finally the saved rates are weighted-averaged according to the number of samples in each subband for the overall rate saving. For H.264/AVC, the prediction results are generated by JM13.2 [13] using the baseline profile, and is wavelet decomposed for comparison.

In the first simulation, we consider the noise-free case, where the frames are finely quantized such that the prediction error is

[2] ρ_Δ can certainly take negative values, but as proved in [11], $\rho_\Delta \geq (1–N)^{-1}$, so here we use 0 as the lower bound of ρ_Δ for large Ns.

Table I. Comparison of the prediction performance between MRMR and H.264/AVC

Saved Rates (bpp) [a]	Akiyo	CarPhone	Container	Football	Foreman	Miss_Am	M&D	News	Suzie	Salesman	Average
MRMR	–2.89	–1.70	–3.48	–0.42	–1.88	–1.46	–2.09	–2.80	–1.10	–2.44	–2.03
H.264/AVC	–2.58	–1.88	–3.04	–0.75	–2.03	–1.53	–2.12	–2.82	–1.31	–2.54	–2.06

[a] Rate saving with a greater absolute value means better prediction.

almost purely due to motion mismatch. The first 100 frames of ten QCIF sequences are tested. The results are listed in Table I. We can see that with extensive motion exploration, the prediction performance of MRMR can be very close to that of state-of-the-art inter-frame coding (the 0.03 bpp difference can be translated into 0.18 dB in PSNR performance), without any overhead in transmitting the motion information. We can also see that typically, if MCP is effective for the sequence, where the achievable rate saving is higher than 2 bpp for both methods, MRMR is as good as, or even better than H.264/AVC; otherwise MRMR still suffers a performance gap(except for the Miss America sequence). This is partly because the intra-prediction mode in H.264/AVC also provides good prediction if no match is found in the ME.

Secondly, we consider the case that the reference frames are noisy. Fig. 4 shows the testing result on the Foreman sequence. We can see that the performance gap becomes smaller for low-quality references. This is because in MHP, the quantization noise from multiple hypotheses is suppressed by averaging.

Finally, in terms of complexity, MRMR also provides some advantage over motion extrapolation, if the same BMA is used in both approaches. For example, when the motion refinement is carried out based on a half-resolution image, the number of MVs for estimation is 1/4 of that of a full-size image. For an *N*-level refinement, the overall complexity is (1/4 + 1/16 + …) < 1/3 of the complexity of the full-resolution motion extrapolation. When compared to the ME module of H.264/AVC, MRMR is certainly more complex, mainly due to the increased number of MVs to determine. However, MRMR is carried out at the decoder side, which makes the encoding process much simpler than H.264/AVC.

5. CONCLUSIONS

Inefficient SI generation is the bottleneck in improving the coding performance of WZVC. Without full access to the current frame, decoder-side ME can never outperform encoder-side ME if the same ME method is employed. However, decoder-side ME can benefit from more detailed motion information, which, if generated and transmitted by the encoder, incurs non-negligible bit overhead. In this paper, based on our previous approach of MRMR, we provide theoretical analysis to the performance achieved by integrating MRMR with advanced ME techniques. The practical SI estimator we implemented shows prediction performance comparable to H.264/AVC.

APPENDIX

Similar to the discussion as in [11], we have

$$\mathbf{P} = P\left(\omega, \sigma_\Delta^2\right)\mathbf{1} \tag{12}$$

and

$$E\left(\mathbf{DD}^H\right) = P\left(\omega, 2(1-\rho_\Delta)\sigma_\Delta^2\right) \times \mathbf{11}^T + diag\left(1 - P\left(\omega, 2(1-\rho_\Delta)\sigma_\Delta^2\right)\right) \cdot \tag{13}$$

where $P\left(\omega,\sigma^2\right) = \exp\left(-\frac{1}{2}\omega^T\omega\sigma^2\right)$ is the Fourier transform of a Gaussian p.d.f., and $\mathbf{1} = [1, \ldots, 1]^T$.

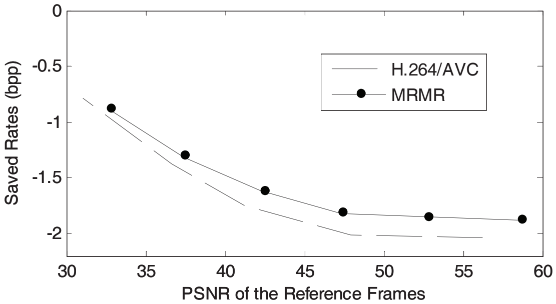

Fig. 4. Rate saving performance with noisy reference frames (Foreman).

On the other hand, if matrix $\mathbf{C} = b\times\mathbf{11}^T + \text{diag}(a - b)$, its inverse can be written as

$$\mathbf{C}^{-1} = \frac{-b\times\mathbf{11}^T + \text{diag}\left(a+(N-1)b\right)}{\left(a-b\right)\left(a-(N-1)b\right)}. \tag{14}$$

Substitute *a* with 1, *b* with $P\left(\omega, 2(1-\rho_\Delta)\sigma_\Delta^2\right)$, and insert (12) – (14) into (8) we get (9).

REFERENCES

[1] W. Liu, L. Dong and W. Zeng, “Wyner-Ziv video coding with multi-resolution motion refinement: Theoretical analysis and practical significance”, (invited paper) *Visual Communications and Image Processing (VCIP)*, San Jose, CA, Jan. 2008.

[2] Z. Li, L. Liu and E. J. Delp, “Rate Distortion Analysis of Motion Side Estimation in Wyner–Ziv Video Coding”, *IEEE Trans. Image Processing*, vol. 16, no. 1, pp. 98–113, Jan. 2007.

[3] A. Aaron, S. Rane and B. Girod, “Wyner-Ziv video coding with hash-based motion compensation at the receiver”, *Proc. IEEE Int. Conf. Image Processing*, Singapore, pp. 3097–3100, Oct. 2004.

[4] R. Puri and K. Ramchandran, “Prism: An uplink-friendly multimedia coding paradigm”, *Proc. IEEE Int. Conf. Image Processing*, Barcelona, Spain, pp. IV 856–859, Sep. 2003.

[5] X. Artigas and L. Torres, “Iterative generation of motion-compensated side information for distributed video coding”, *Proc. IEEE Int. Conf. Image Processing*, Genova, Italy, pp. IV 833–836, Sep. 2005.

[6] J. Ascenso, C. Brites, and F. Pereira, “Motion compensated refinement for low complexity pixel based distributed video coding”, *Proc. IEEE Int. Conf. Advanced Video and Signal-Based Surveillance*, Como, Italy, pp. 593–598, Sep. 2005.

[7] B. Girod, “The efficiency of motion-compensating prediction for hybrid coding of video sequences,” *IEEE J. Sel. Areas. Commun.*, vol. 5, no. 8, pp. 1140–1154, Aug. 1987.

[8] —, “Motion-compensating prediction with fractional-pel accuracy,” *IEEE Trans. Commun.*, vol. 41, no. 4, pp. 604–612, Apr. 1993.

[9] —, “Efficiency analysis of multihypothesis motion-compensated prediction for video coding”, *IEEE Trans. Image Proc.*, vol. 9, no. 2, pp. 173–183, Feb. 2000.

[10] R. Buschmann, “Efficiency of displacement estimation techniques”, *Signal Processing: Image Communication*, vol. 10, pp. 43–61, 1997.

[11] M. Flierl, T. Wiegand, and B. Girod, “Rate-constrained multihypothesis prediction for motion-compensated video compression,” *IEEE Trans. Circ. Sys. Video Tech.*, vol. 12, no. 11, pp. 957–969, Nov. 2002.

[12] Draft ITU-T Recommendation and Final Draft International Standard, Pattaya, Thailand, 2003.

[13] Available online http://iphome.hhi.de/suehring/tml/.