# BLOCK SIZE DEPENDENT ERROR MODEL FOR MOTION COMPENSATION

*Sven Klomp, Marco Munderloh, Jörn Ostermann, Fellow, IEEE*

Institut für Informationsverarbeitung
Leibniz Universität Hannover, Appelstr. 9A, 30167 Hannover, Germany
{klomp, munderl, ostermann}@tnt.uni-hannover.de

## ABSTRACT

Current video coding standards use block-based motion estimation and compensation algorithms to exploit dependencies between consecutive frames. It is a well-known fact that decreasing the block size reduces the motion-compensated frame difference, and thus reduces the data rate. However, no theoretical evaluations are available to model this relation.

This paper derives a model for the prediction error variance of block-based motion compensation algorithms with respect to the block size. It is shown that the variance of the displaced frame difference of a block can be modelled with the pixel position and only three additional parameters. It can be observed that the variance increases almost linearly with the block size.

***Index Terms***— Video coding, motion compensation, block size, block matching, prediction error

## 1. INTRODUCTION

Although a lot of motion estimation techniques, like optical flow [1] or mesh-based [2] algorithms, were developed and improved within the last years, the major video coding standards, such as MPEG-1,2,4 video or ITU-T H.26x, use block-based algorithms for performance and implementation reasons. The block size in those algorithms highly affects the quality of the predicted frame [3]. Large blocks can contain several objects moving in different directions and thus, the motion compensation fails. Smaller blocks can better adapt to local motion, and would therefore result in a more accurate prediction. However, the smaller the block size, the more motion vectors have to be coded and transmitted to the receiver. Therefore, the block size is chosen by a rate distortion optimisation, where both the rate for the prediction error and the rate for the motion vectors are taken into account.

Miscellaneous characteristics of the motion-compensated frame difference are evaluated in the literature. In [4], the frame to frame difference is empirically determined. The fact that the prediction error caused by the motion compensation is not homogeneous within a block is described in [5]. Figure 1 shows the mean prediction error of each pixel for a part of the Kimono sequence which is notably larger at the block boundaries.

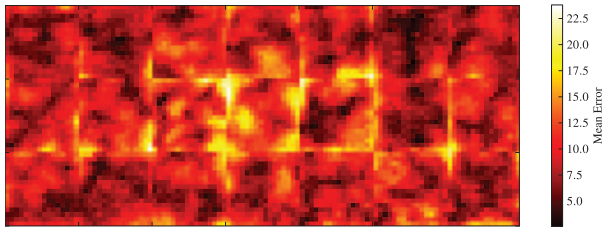


**Fig. 1**. Detail of the prediction error for the Kimono sequence

However, the impact of the block size is not considered in those evaluations. The relationship between the motion-compensated frame difference and the block size allows for a better understanding of current video coding standards. Furthermore, the evaluation of small block sizes is of special interest for methods like decoder-side motion estimation [6], as no motion vectors are transmitted, and thus, no lower limit for the block size exists.

Therefore, this paper proposes a model for motion-compensated frame difference in Section 2, which takes the block size into account. Section 3 shows experimental results, and the paper finishes with conclusions in Section 4.

## 2. MODEL OF MOTION-COMPENSATED PREDICTION ERROR

It is assumed that the image intensities of two consecutive frames $f_t(x,y)$ and $f_{t+1}(x,y)$ are linked by

$$f_t(x,y) = f_{t+1}(x + d_x(x,y), y + d_y(x,y)) + n_0(x,y) \quad (1)$$

where $\vec{d}(x,y) = (d_x(x,y), d_y(x,y))$ is the motion vector field between the two frames representing the true motion for each pixel. To incorporate errors caused by non-translational motion and occlusion, a zero-mean noise variable $n_0(x,y)$ with variance $\sigma^2_{n_0}$ is added.

In the following examination, the prediction error for a $M \times N$ block within frame $f_t$ is derived. For simplification, the coordinate origin is set to the middle of the block as shown in Figure 2.

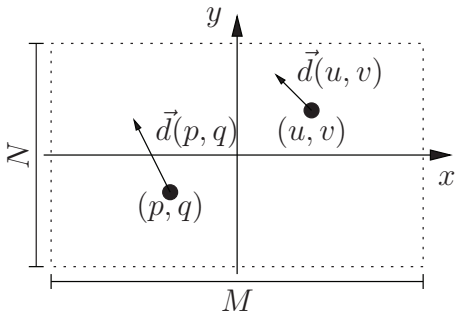


**Fig. 2**. A block of size $M \times N$ in the image $f_t$

It is assumed that the motion vector obtained by the block-based motion estimation algorithm equals the vector at position $(p,q)$. Hence, the estimated motion vector is an element of the set of true motion vectors of the particular block. If this vector is used for motion compensation, the displaced difference for a pixel position $(u,v)$ within this block results in the following equation and can be approximated by a first order Taylor polynomial:

$$\begin{aligned} & f_t(u,v) - f_{t+1}(u + d_x(p,q), v + d_y(p,q)) \\ = \;& f_{t+1}(u + d_x(u,v), v + d_y(u,v)) + n_0(u,v) \\ & - f_{t+1}(u + d_x(p,q), v + d_y(p,q)) \qquad (2) \\ \overset{Taylor}{=} \;& (d_x(u,v) - d_x(p,q)) \cdot \\ & \frac{\delta}{\delta x} f_{t+1}(u + d_x(p,q), v + d_y(p,q)) \\ & + (d_y(u,v) - d_y(p,q)) \cdot \\ & \frac{\delta}{\delta y} f_{t+1}(u + d_x(p,q), v + d_y(p,q)) \\ & + n_T(\ldots) + n_0(u,v) \qquad (3) \\ = \;& n_e(u,v) \end{aligned}$$

with $n_T(\ldots)$ specifying the error caused by the first order Taylor approximation. This error term depends on the difference $\vec{d}(u,v) - \vec{d}(p,q)$ and the image gradient $(\frac{\delta}{\delta x} f_{t+1}, \frac{\delta}{\delta y} f_{t+1})$. For simplification, it is assumed that $n_T(\ldots)$ is proportional to the first two addends of Equation (3). The experimental results in Section 3 show that this simplification is appropriate. Thus, a block size dependent factor $k$ is introduced for the Taylor polynomial, and Equation 3 results in

$$\begin{aligned} n_e(u,v) = k\Big[ & (d_x(u,v) - d_x(p,q)) \\ & \frac{\delta}{\delta x} f_{t+1}(u + d_x(p,q), v + d_y(p,q)) \\ & + (d_y(u,v) - d_y(p,q)) \\ & \frac{\delta}{\delta y} f_{t+1}(u + d_x(p,q), v + d_y(p,q)) \Big] \\ & + n_0(u,v). \qquad (4) \end{aligned}$$

In [5], a statistical model for the relation of neighbouring motion vectors is introduced:

$$d_x(u,v) - d_x(p,q) \sim N\big(0, c_H^2 \left((u-p)^2 + (v-q)^2\right)\big) \quad (5)$$

$$d_y(u,v) - d_y(p,q) \sim N\big(0, c_V^2 \left((u-p)^2 + (v-q)^2\right)\big) \quad (6)$$

where $c_H$ and $c_V$ are constants representing the amount of motion changes in horizontal and vertical directions, respectively. It was shown in [5] that block matching using the squared sum of differences (SSD) will result in motion vectors that are most likely to be the motion vectors at block centres ($E[p] = E[q] = 0$).

However, to get the prediction error to depend on the block size, a closer look on the distribution of $p$ and $q$ is needed. The probability distributions of the two variables are considered Gaussian with block size dependent variances $\sigma_p^2$ and $\sigma_q^2$:

$$p \sim N(0, \sigma_p^2) \qquad (7)$$

$$q \sim N(0, \sigma_q^2) \qquad (8)$$

Evaluations with the synthetic sequences Yosemite (global motion) and Street (global and local motion), for which the motion vectors are known for each pixel, have shown that the probabilities $P(|p| < M/2)$ and $P(|q| < N/2)$ are independent of the block size. In other words, the probability that the estimated motion vector equals a vector within the block is nearly constant:

| **Sequence** | **4x4** | **8x8** | **16x16** |
|---|---|---|---|
| **Yosemite** | 92% | 88% | 91% |
| **Street** | 98% | 94% | 99% |

Although the accuracy of the motion estimation decreases for larger block sizes, the amount of possible candidates increases and the overall probability does not change significantly.

Therefore, the variances of the probability density functions for $p$ and $q$ are proportional to the block size, as shown for $p$ in Figure 3. Thus, the Equations (7) and (8) can be written as

$$p \sim N(0, M^2 \sigma_{p0}^2) \qquad (9)$$

$$q \sim N(0, N^2 \sigma_{q0}^2) \qquad (10)$$

where $\sigma_{p0}^2$ and $\sigma_{q0}^2$ are variances independent of the block size.

After all parameters in Equation (4) are specified, it is possible to calculate the variance of the motion-compensated frame difference for each pixel position within a block:

$$\sigma_e^2(u,v) = E\left[n_e^2(u,v)\right] - E^2\left[n_e(u,v)\right] \qquad (11)$$

The mean of $n_e(u,v)$ is zero, since it is assumed that the gradients $\frac{\delta}{\delta x} f_{t+1}$ and $\frac{\delta}{\delta y} f_{t+1}$ are statistically independent of the

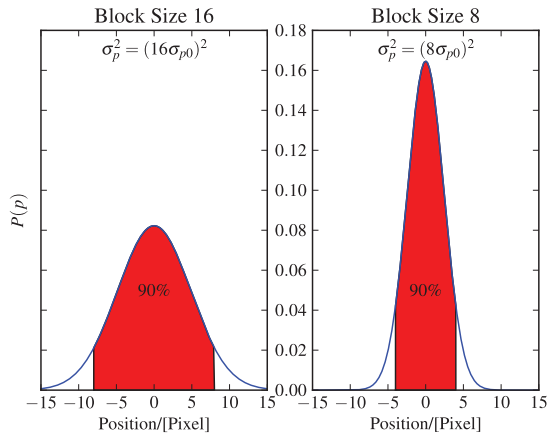


**Fig. 3**. Example of the probability distributions of $p$ for different block sizes.

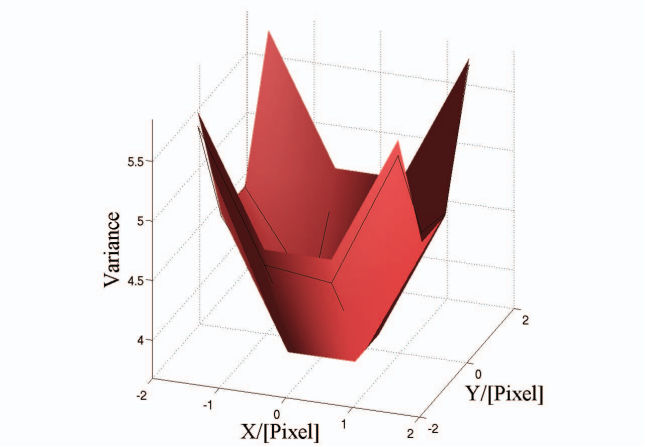


**Fig. 4**. Fitted function (red) and measured data (black grid) for the Kimono sequence with a block size of $4 \times 4$.

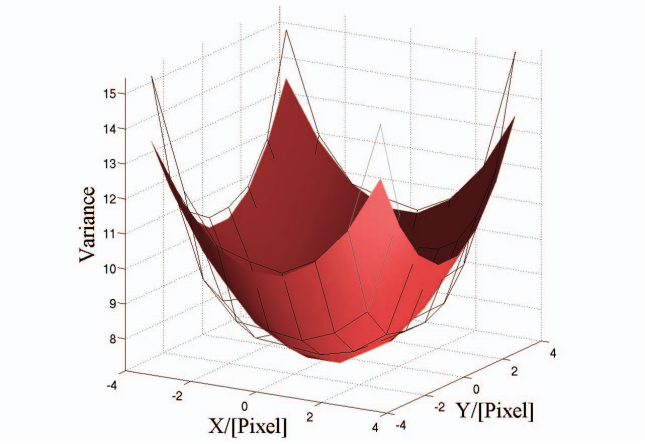


**Fig. 5**. Fitted function (red) and measured data (black grid) for the Kimono sequence with a block size of $8 \times 8$.

motion model (Equation (5) and (6)) and the noise term $\sigma_{n_0}^2$. Furthermore, the second moment $E[(\ldots)^2]$ can be calculated independently for each component of Equation 4 because of these assumptions.

In contrast to [5], the variances of the motion vector differences $d_x(u,v) - d_x(p,q)$ and $d_y(u,v) - d_y(p,q)$ depend on the statistics of the variables $p$ and $q$ due to the assumptions of Equations (9) and (10). The second moments of the differences $d_x(u,v) - d_x(p,q)$ and $d_y(u,v) - d_y(p,q)$ lead to

$$E\left[\left(d_x(u,v)-d_x(p,q)\right)^2\right] = c_H^2\left(u^2+\sigma_p^2+v^2+\sigma_q^2\right) \quad (12)$$

$$E\left[\left(d_y(u,v)-d_y(p,q)\right)^2\right] = c_V^2\left(u^2+\sigma_p^2+v^2+\sigma_q^2\right) \quad (13)$$

and thus, the variance results in

$$\begin{aligned}\sigma_e^2(u,v) = k\Big[ & c_H^2\left(u^2+\sigma_p^2+v^2+\sigma_q^2\right)\sigma_H^2 \\ & + c_V^2\left(u^2+\sigma_p^2+v^2+\sigma_q^2\right)\sigma_V^2\Big] \\ & + \sigma_{n_0}^2 \end{aligned} \quad (14)$$

where $\sigma_H^2$ and $\sigma_V^2$ are the variances of the image gradients:

$$\sigma_H^2 = E\left[\left(\frac{\delta}{\delta x} f_{t+1}(x,y)\right)^2\right] \quad (15)$$

$$\sigma_V^2 = E\left[\left(\frac{\delta}{\delta y} f_{t+1}(x,y)\right)^2\right] \quad (16)$$

For quadratic blocks $M \times M$, Equation (14) can be modelled with three independent parameters $A$, $B$ and $\sigma_{n_0}^2$, the block size $M$ and the pixel position $(m,n)$ within a block:

$$\sigma_e^2(u,v) = kA\left(u^2+v^2\right) + kAM^2B + \sigma_{n_0}^2 \quad (17)$$

with the constants $A = \left(c_H^2\sigma_H^2 + c_V^2\sigma_V^2\right)$ and $B = \sigma_{p0}^2 + \sigma_{q0}^2$.

## 3. EXPERIMENTAL VERIFICATION

This section deals with the verification of the model proposed in the previous section. A conventional block-based motion estimation algorithm with half-pel accuracy which minimises the sum of squared differences (SSD) is used to calculate a prediction of the current frame. This prediction is subtracted from the original frame, yielding the prediction error. For each pixel in a $M \times M$ block, the variance of the prediction error is calculated over all blocks of the sequence. To fit Equation (17) to the measured data, the least squares method is used. For the HD sequence Kimono, the fitting for different block sizes are shown in Figures 4, 5 and 6. Other test sequences give similar results.

The model in red approximates the measured data in black very well. However, the accuracy slightly decreases for larger blocks, since the statistical motion model (Equation (5)

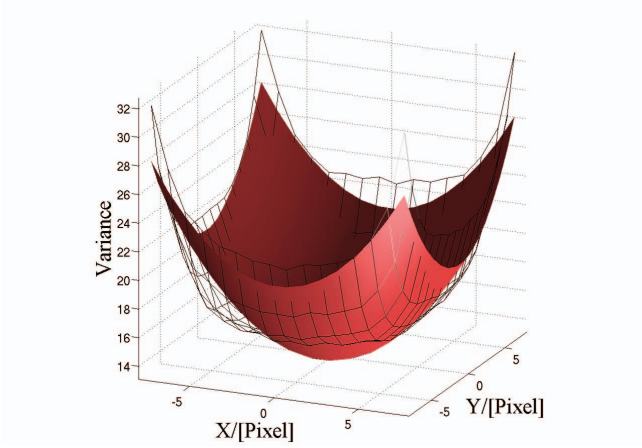


**Fig. 6**. Fitted function (red) and measured data (black grid) for the Kimono sequence with a block size of $16 \times 16$.

| Sequence | M | $kA$ | $B$ | $\sigma_{n_0}^2$ | FE(%) |
|---|---|---|---|---|---|
| **Kimono** | 4 | 0.418 | 0.343 | 1.356 | 2.12 |
| | 8 | 0.261 | 0.343 | 1.356 | 4.42 |
| | 16 | 0.134 | 0.343 | 1.356 | 5.75 |
| **People on Street** | 4 | 2.679 | 0.324 | 3.801 | 3.38 |
| | 8 | 1.259 | 0.324 | 3.801 | 5.20 |
| | 16 | 0.615 | 0.324 | 3.801 | 6.40 |
| **Foreman** | 4 | 0.385 | 0.989 | 0.692 | 3.93 |
| | 8 | 0.159 | 0.989 | 0.692 | 5.18 |
| | 16 | 0.063 | 0.989 | 0.692 | 9.86 |

**Table 1**. Results of data fitting.

and (6)) is only appropriate for small coordinate differences.

To get a quantitative evaluation of the fitting accuracy, the fitting error

$$\mathrm{FE} = \left[ \frac{1}{M^2} \sum_{u=-\frac{M-1}{2}}^{\frac{M-1}{2}} \sum_{u=-\frac{M-1}{2}}^{\frac{M-1}{2}} \frac{|\sigma_{e,1}^2(u,v) - \sigma_{e,2}^2(u,v)|}{\sigma_{e,2}^2(u,v)} \right] \cdot 100\% \tag{18}$$

as proposed in [5], is computed, where $\sigma_{e,1}^2(u,v)$ is the measured variance and $\sigma_{e,2}^2(u,v)$ is the calculated variance of the displaced frame difference. The results are shown in Table 1. Kimono and People on Street are HD sequences with $1920 \times 1080$ and $2560 \times 1600$ pixel resolution, respectively. The size of the Foreman sequence is CIF.

The fitting error FE increases with the block size, since the motion model is more accurate for small intervals as previously mentioned and fitting with less points can lower the fitting error. For the Foreman sequence and a block size of $16 \times 16$, FE increases significantly due to the small image resolution.

The parameters $A$, $B$ and $\sigma_{n_0}^2$ depend only on the motion and frame content and thus, only $k$ depends on the block size. Interestingly, $k$ is proportional to $\frac{1}{M}$ and doubles for halved blocks. Therefore, Equation 17 can be written as

$$\sigma_e^2(u,v) = \frac{A}{M}\left(u^2+v^2\right) + AMB + \sigma_{n_0}^2 \tag{19}$$

and it can be noticed that the error variance in the middle of the block increases linearly with larger block sizes.

The error variance $\sigma_{n_0}^2$ caused by non-translational motion and occlusion is higher for the People on Street sequence, where several people are crossing a street. The Kimono sequence contains a camera pan and only one moving person, and thus has less occlusion.

## 4. CONCLUSIONS

In this paper, a model to calculate the error variance for block-based motion compensation is proposed. Only three sequence-dependent parameters are needed to get an accurate approximation of the motion-compensated frame difference. This analysis gives an insight into the relation between the block size used during motion estimation and the variance of the displaced frame difference. It was observed that the variance at the block centre increases in a linear way with respect to the block size.

## 5. REFERENCES

[1] T. Brox, A. Bruhn, N. Papenberg, and Joachim Weickert, "High accuracy optical flow estimation based on a theory for warping," in *Proc. ECCV*, Prague, Czech Republic, May 2004, vol. 4, pp. 25–36.

[2] Y. Nakaya and H. Harashima, "Motion compensation based on spatial transformations," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 4, no. 3, pp. 339–356, June 1994.

[3] M. Wien, "Variable block-size transforms for h.264/avc," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 604–613, July 2003.

[4] D.J. Connor and J.O. Limb, "Properties of frame-difference signals generated by moving images," *IEEE Trans. Commun.*, vol. 22, no. 10, pp. 1564–1575, October 1974.

[5] W. Zheng, Y. Shishikui, M. Naemura, Y. Kanatsugu, and Susumu Itoh, "Analysis of space-dependent characteristics of motion-compensated frame differences based on a statistical motion distribution model," *IEEE Trans. Image Processing*, vol. 11, no. 4, pp. 377–386, April 2002.

[6] S. Klomp, M. Munderloh, Y. Vatis, and J. Ostermann, "Decoder-side block motion estimation for h.264 / mpeg-4 avc based video coding," in *Proc. ISCAS*, Taipei, Taiwan, May 2009, pp. 1641–1644.