Reprinted from

IMAGE COMMUNICATION

Signal Processing: *Image Communication* 10 (1997) 43–61

Efficiency of displacement estimation techniques

Ralf Buschmann*

Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, Universität Hannover, Appelstraße 9A, D-30167 Hannover, Germany

ELSEVIER

IMAGE COMMUNICATION

Theory, Techniques & Applications

A publication of the European Association for Signal Processing (EURASIP)

Editorial Policy. SIGNAL PROCESSING: IMAGE COMMUNICATION is an international journal for the development of the theory and practice of image communication. Its primary objectives are the following:

- To present a forum for the advancement of the theory and practice of image communication.
- To stimulate cross-fertilization between areas similar in nature which have traditionally been separated, for example, various aspects of visual communications and information systems.
- To contribute to a rapid information exchange between the industrial and academic environments.

The editorial policy and the technical content of the journal are the responsibility of the Editor-in-Chief and the Editorial Board. The Journal is self-supporting from subscription income and contains a minimum amount of advertisements. Advertisements are subject to the prior approval of the Editor-in-Chief. The journal welcomes contributions from every country in the world.

Scope. SIGNAL PROCESSING: IMAGE COMMUNICATION publishes articles relating to aspects of the design, implementation and use of image communication systems.
SIGNAL PROCESSING: IMAGE COMMUNICATION features original research work, tutorial and review articles, and accounts of practical developments.

Subjects. Subject areas covered by the journal include:
TV, HDTV and 3D-TV systems
Visual Science
Image Coding
TV and Advanced TV Broadcasting
Image Storage and Retrieval
Graphic Arts
Electronic Printing
Image Transmission
Interactive Image Communication
Imaging Technology
Display Technology
VLSI Processors for Image Communications

Membership and Subscription Information.
SIGNAL PROCESSING: IMAGE COMMUNICATION (ISBN 0923-5965) is published in two volume (eight issues) a year. For 1997 Volumes 9 and 10 are scheduled for publication. Subscription prices are available upon request from the publisher. Subscriptions are accepted on a prepaid basis only and are entered on a calendar year basis. Issues are sent by surface mail except to the following countries, where air delivery (S.A.L. – Surface Air Lifted) is ensured: Argentina, Australia, Brazil, Canada, Hong Kong, India, Israel, Japan, Malaysia, Mexico, New Zealand, Pakistan, People's Republic of China, Singapore, South Africa, South Korea, Taiwan, Thailand, USA. For the rest of the world, airmail charges are available upon request. Claims for missing issues will be honoured free of charge if made within six months after the publication date of the issues. For orders, claims, product enquiries (no manuscript enquiries) please contact the Customer Support Department at the Regional Sales Office nearest to you:
New York, Elsevier Science, P.O. Box 945, New York, NY 10159-0945, USA. Tel: (+1) 212-633-3730, [Toll Free number for North American Customers: 1-888-4ES-INFO (437-4636)], Fax: (+1) 212-633-3680, E-mail: usinfo-f@elsevier.com
Amsterdam, Elsevier Science, P.O. Box 211, 1000 AE Amsterdam, The Netherlands. Tel: (+31) 20-485-3757, Fax: (+31) 20-485-3432, E-mail: nlinfo-f@elsevier.nl
Tokyo, Elsevier Science, 9-15, Highashi-Azabu 1-chome, Minato-ku, Tokyo 106, Japan. Tel: (+81) 3-5561-5033, Fax: (+81) 3-5561-5047, E-mail: kyf04035@niftyserve.or.jp
Singapore, Elsevier Science, No. 1 Temasek Avenue, #17-01 Millenia Tower, Singapore 039192. Tel: (+65) 434-3727, Fax: (+65) 337-2230, E-mail: asiainfo@elsevier.com.sg
For full membership information of the Association, possibly combined with a subscription at a reduced rate, please contact: EURASIP, P.O. Box 134, CH-1000 Lausanne 13, Switzerland.

Published 8 times a year 0165-1684/97/$17.00 Printed in The Netherlands

ELSEVIER

Signal Processing: *Image Communication* 10 (1997) 43–61

SIGNAL PROCESSING: IMAGE COMMUNICATION

Efficiency of displacement estimation techniques

Ralf Buschmann*

Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, Universität Hannover, Appelstraße 9A, D-30167 Hannover, Germany

Abstract

An analytical description of displacement estimation which allows an objective evaluation and comparison of various displacement estimation techniques is presented. For evaluation and comparison rate-distortion functions are calculated and by this the impact of the 2D motion model of the displacement estimator and the amplitude and spatial resolution of the estimated sparse displacement vector field can be quantified. The rate-distortion function describes the relationship between the encoding bit-rate required for transmission of the measured displacement vector field and the variance of the displacement error. The impact of the spatial and amplitude resolution of the sparse displacement vector field is described by sampling and quantisation of an exact displacement signal. The 2D motion model is considered by low pass filtering the exact displacement signal before sampling and subsequent spatial interpolation. The analytical description is verified by simulations of various displacement estimation techniques, the 2D motion models of which can be described by the nearest neighbour, affine and bilinear displacement vector interpolation, at different amplitude and spatial resolutions. The analytical description allows one to optimise a displacement estimation technique for motion compensated image sequence coding. © 1997 Elsevier Science B.V.

Keywords: Image coding; Generalized displacement estimation; Displacement coding; Rate distortion theory

1. Introduction

For efficient coding of moving images block-based motion compensated coding is applied in todays coding standards. The motion is represented by displacement vector fields, which are measured with help of displacement estimation techniques. In order to improve the efficiency of block-based motion compensated coding

- displacement estimation with variable displacement amplitude resolution [5],
- displacement estimation with variable spatial resolution [2, 10],
- displacement estimation using 2D transformations [3, 11, 13, 15, 16, 19]

have been proposed in the recent literature and are currently investigated in MPEG-4 [6, 7].

In displacement estimation with variable amplitude resolution, the amplitude resolution of the displacement vector is not kept fixed at integer or half pel resolution as in todays standards, but can vary adaptively among various resolutions, e.g. 0.5, 0.25 or 0.125 pel.

In displacement estimation with variable spatial resolution the spatial distance of displacement

*Tel.: + 49 (0) 511 762 5308; fax: + 49 (0) 511 762 5333; e-mail: buschman@tnt.uni-hannover.de.

PII S0923-5965(97)00018-0

vectors varies adaptively, e.g. among 4×4, 8×8, 16×16, 32×32 and 64×64. Thus the displacement vector field can be represented with increased spatial resolution near the boundaries of moving objects.

Displacement estimation techniques using 2D transformations can be described by sparse displacement estimation combined with subsequent spatial interpolation. Although the subsequent spatial interpolation is not always carried out explicitly, it is inherent in the 2D transformation. Typical interpolation techniques applied are the affine or bilinear interpolation of estimated displacement vectors at node locations of a regular triangular or quadrilateral mesh overlayed on the image. The attachment of a single estimated displacement vector to a whole block in traditional block matching can be viewed as nearest neighbour interpolation of displacement vectors at node locations of a regular quadrilateral mesh.

The aim of this work is an objective assessment of the efficiency of the displacement estimation techniques in a video coding system by their rate-distortion functions. The rate-distortion function of a displacement estimation technique describes the relationship between the theoretical lower bound of its displacement estimation error and the encoding bit-rate required for transmission of an estimated sparse displacement vector field.

Both, the theoretical lower bound of the displacement estimation error and the encoding bit-rate of the estimated sparse displacement vector field, depend on its amplitude and spatial resolution. The theoretical lower bound of the displacement estimation error additionally depends on the kind of displacement vector interpolation. An objective assessment of the various estimation techniques independent of any test sequence and any implemented optimisation strategy is only possible with an analytical description of displacement estimation.

For this purpose, in its first part, this paper describes displacement estimation analytically giving additional insight into the sources of displacement errors. In the second part, rate-distortion theory is combined with the analytical description of displacement estimation to appraise the encoding bit-rate for the displacement vector fields.

The rate-distortion functions of displacement estimation techniques presented in this paper relate the lower bound of the displacement estimation error to the encoding bit-rate of the displacement information. An analytical description that relates the variance of the displacement estimation error to the encoding bit-rate of the prediction error image needed to gain a predefined image quality is presented in [4]. Combining the results of the work presented in this paper with the work presented in [4], is seen as a good start to reach the goal of finding the theoretically optimum bit allocation for the displacement information and the prediction error image.

In Section 2 the analytical description of displacement estimation is derived. Section 3 extends the analytical description by consideration of displacement coding. In Section 4 the model parameters for the exact displacement signal are given. Experimental results are presented and discussed in Section 5. Section 6 contains the conclusions.

2. Analytical description of displacement estimation

In this work the exact displacement signal is assumed to be known. It is viewed as a two-dimensional signal, with two components, that exists in the image plane. It cannot be measured directly, but it leads to the temporal changes observable in the image sequence. Each component of the displacement signal is assumed to be discrete in time, but space- and amplitude-continuous (the amplitude is real valued). Discontinuities in their amplitude may occur at the boundaries of moving objects. Effects of undefined areas in the displacement signal, e.g. in uncovered background regions are not considered in this work.

With this assumption displacement estimation is described analytically by low pass filtering, sampling, and quantisation of an exact displacement signal combined with subsequent spatial interpolation. The displacement estimation error is then the difference between the exact and the reconstructed displacement signal (Fig. 1). Using the power spectral density of the exact displacement signal as a reference, with this analytical description the

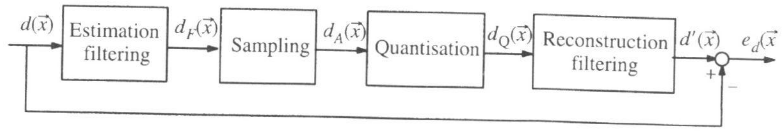


Fig. 1. Block diagram of filtering, sampling, quantisation and reconstruction of a displacement signal.

displacement error variance of the various displacement estimation techniques can be determined.

The low pass filtering of the displacement signal before sampling considers the fact that during displacement estimation a local neighbourhood in the image signal is used to determine one displacement vector. Thus, if the picture elements in the local neighbourhood have been displaced differently the estimated displacement vector is not a sample of the displacement signal, but is an averaged value of the displacements in the local neighbourhood. The frequency response of this estimation filter depends on the 2D transformation model of the displacement estimator.

The sampling of this filtered displacement signal reduces its spatial resolution. Different sampling rates correspond to different spatial distances for which a displacement vector is estimated. Thus it is possible to describe e.g. block matching with different block sizes.

Subsequent quantisation of the samples introduces variable amplitude resolution of the estimated displacement vectors.

By displacement vector interpolation a reconstruction of the displacement signal is generated. The vector interpolation is inherent in the applied 2D transformation of the displacement estimator. Like the frequency response of the estimation filter, the frequency response of the reconstruction filter depends on the 2D transformation model of the displacement estimator. They allow the analysis of different 2D transformation models. Interpolation techniques examined in this work are the affine and bilinear interpolation of estimated displacement vectors at node locations of a regular triangular or quadrilateral mesh. The attachment of a single estimated displacement vector to a whole block in traditional block matching is viewed as nearest neighbour interpolation of displacement vectors at node locations of a regular quadrilateral mesh.

The examined interpolation techniques require that both the estimation filter and the reconstruction filter are applied separately to the two components of the displacement signals. Thus for simplicity of the following description all signals are viewed as one-component signals.

The aim of the following is to relate the power spectral density $S_{dd}(f_x, f_y)$ of the exact displacement signal $d(x, y)$ to the power spectral density $S_{e_d e_d}(f_x, f_y)$ of the displacement error signal $e_d(x, y)$. The exact displacement signal $d(x, y)$ is assumed to be a real zero-mean ergodic wide-sense stationary stochastic process. It is important to note here that due to the sampling, the sampled displacement signal $d_A(x, y)$, the quantised displacement signal $d_Q(x, y)$, the reconstructed displacement signal $d'(x, y)$, and thus the displacement error signal, are no longer wide-sense stationary. However, they can be described as wide-sense cyclostationary, which means that their mean and autocorrelation fluctuate periodically in time. In other words, their mean and autocorrelation are invariant to a shift of the origin by integral multiples of $T_{d,x}$ in the x-direction and/or of $T_{d,y}$ in the y-direction, where $T_{d,x}$ and $T_{d,y}$ denote the horizontal and vertical sampling periods. Cyclostationary processes may be treated as they were stationary, by simply averaging the statistical parameters over one cycle [18].

To reach the goal of relating the power spectral density of the exact displacement signal to the power spectral density of the displacement error signal, in a first step the estimation filtering, the sampling, the quantisation, and the reconstruction filtering are described separately. Combining these descriptions in a second step, the desired relation is derived.

2.1. Estimation filtering

Let $H_1(f_x, f_y)$ denote the frequency response of the estimation filter. The power spectral density

$S_{d_F d_F}(f_x, f_y)$ of the filtered displacement signal, denoted by $d_F(x, y)$, is then given by

$$S_{d_F d_F}(f_x, f_y) = S_{dd}(f_x, f_y)|H_1(f_x, f_y)|^2. \quad (1)$$

2.2. Sampling

The power spectral density $S_{d_A d_A}(f_x, f_y)$ of the sampled displacement signal, denoted by $d_A(x, y)$, is given by normalised shifted copies of the power spectral density of the filtered displacement signal [18]:

$$S_{d_A d_A}(f_x, f_y) = \frac{1}{T_{d,x} T_{d,y}} \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} S_{d_F d_F} \times \left(f_x + \frac{m}{T_{d,x}}, f_y + \frac{n}{T_{d,y}} \right), \quad (2)$$

with $T_{d,x}$ and $T_{d,y}$ denoting the horizontal and vertical sampling periods. For the simplicity of further writing let us simplify (2) by

$$S_{d_A d_A}(f_x, f_y) = \frac{1}{T_{d,x} T_{d,y}} (S_{d_F d_F}(f_x, f_y) + S_{n_A n_A}(f_x, f_y)), \quad (3)$$

with

$$S_{n_A n_A}(f_x, f_y) = \sum_{\substack{n=-\infty \\ (m,n) \neq (0,0)}}^{\infty} \sum_{m=-\infty}^{\infty} S_{d_F d_F}\left(f_x + \frac{m}{T_{d,x}}, f_y + \frac{n}{T_{d,y}} \right)$$
$$= \sum_{\substack{n=-\infty \\ (m,n) \neq (0,0)}}^{\infty} \sum_{m=-\infty}^{\infty} S_{dd}\left(f_x + \frac{m}{T_{d,x}}, f_y + \frac{n}{T_{d,y}} \right) \times \left| H_1\left(f_x + \frac{m}{T_{d,x}}, f_y + \frac{n}{T_{d,y}} \right) \right|^2, \quad (4)$$

describing the sum of shifted copies of the power spectral density of the filtered displacement signal, which may lead to aliasing noise in the reconstructed displacement signal.

2.3. Quantisation

The influence of the quantiser is described by a quantisation noise signal denoted by $n_Q(x, y)$. The quantisation noise signal is assumed to be independent of the sampled displacement signal, which is valid when the quantisation step size is small with respect to the variance of the sampled displacement signal. The power spectral density $S_{n_Q n_Q}(f_x, f_y)$ of the quantisation noise signal is assumed to be white with variance $\sigma^2_{n_Q}$:

$$S_{n_Q n_Q}(f_x, f_y) = \sigma^2_{n_Q}. \quad (5)$$

Assuming uniform quantisation of $d(\boldsymbol{x})$ with step size Δ and K quantisation steps and a probability density function $p_d(d)$ of the displacement signal the quantisation error variance can be calculated by

$$\sigma^2_{n_Q} = \sum_{k=1}^{K} \int_{d_{Q,k}-\Delta/2}^{d_{Q,k}+\Delta/2} (d_{Q,k} - d)^2 p_d(d)\, \mathrm{d}d. \quad (6)$$

The power spectral density $S_{n_Q n_Q}(f_x, f_y)$ of the quantised displacement signal is then given by

$$S_{d_Q d_Q}(f_x, f_y) = S_{d_A d_A}(f_x, f_y) + S_{n_Q n_Q}(f_x, f_y). \quad (7)$$

2.4. Reconstruction filtering

Considering the cyclostationarity of the reconstructed displacement signal, denoted by $d'(x, y)$, its averaged power spectral density $\bar{S}_{d'd'}(f_x, f_y)$ is given by [18]

$$\bar{S}_{d'd'}(f_x, f_y) = \frac{1}{T_{d,x} T_{d,y}} S_{d_Q d_Q}(f_x, f_y)|H_2(f_x, f_y)|^2, \quad (8)$$

where $H_2(f_x, f_y)$ denotes the frequency response of the reconstruction filter.

2.5. Power spectral density of the displacement error signal

With the given separate descriptions, it is now straightforward to relate the power spectral density of the displacement error signal to the power spectral density of the displacement signal. The autocorrelation function $R_{e_d e_d}(\tau_x, \tau_y)$ of the cyclostationary displacement error signal,

$$R_{e_d e_d}(\tau_x, \tau_y) = E[e_d(x, y) e_d(x - \tau_x, y - \tau_y)]$$
$$= E[(d'(x, y) - d(x, y))(d'(x - \tau_x, y - \tau_y) - d(x - \tau_x, y - \tau_y))]$$
$$= R_{dd}(\tau_x, \tau_y) - R_{dd'}(\tau_x, \tau_y) - R_{dd'}(-\tau_x, -\tau_y) + R_{d'd'}(\tau_x, \tau_y), \quad (9)$$

depends on the spatial reference given by the sampling points. $R_{dd'}(\tau_x, \tau_y) = R_{dd'}(\tau_x, \tau_y)$ denote the cross-correlation between the exact displacement signal and the reconstructed displacement signal. We can treat the displacement error signal as stationary, by averaging over one cycle. The power spectral density $S_{e_d e_d}(f_x, f_y)$ of the displacement error signal is then given by

$$S_{e_d e_d}(\boldsymbol{f}) = S_{dd}(\boldsymbol{f}) - 2\bar{S}_{dd'}(\boldsymbol{f}) + \bar{S}_{d'd'}(\boldsymbol{f}),$$
$$\text{with } \boldsymbol{f} = (f_x, f_y)^{\mathrm{T}}, \quad (10)$$

where $\bar{S}_{dd'}(f_x, f_y)$ denotes the averaged cross-power spectral density between the exact displacement signal and the reconstructed displacement signal. In (10) it has been considered that the power spectral densities and frequency responses examined in this work are real.

Using Eqs. (1)–(8) the averaged power spectral density of the reconstructed displacement signal follows:

$$\bar{S}_{d'd'}(\boldsymbol{f}) = \frac{1}{T_{d,x} T_{d,y}} S_{d_Q d_Q}(\boldsymbol{f})|H_2(\boldsymbol{f})|^2$$
$$= \frac{1}{T_{d,x} T_{d,y}} (S_{d_A d_A}(\boldsymbol{f}) + S_{n_Q n_Q}(\boldsymbol{f}))|H_2(\boldsymbol{f})|^2$$
$$= \frac{1}{(T_{d,x} T_{d,y})^2} S_{dd}(\boldsymbol{f})|H_1(\boldsymbol{f}) H_2(\boldsymbol{f})|^2 + \frac{1}{T_{d,x} T_{d,y}} S_{n_Q n_Q}(\boldsymbol{f})|H_2(\boldsymbol{f})|^2 + \frac{1}{(T_{d,x} T_{d,y})^2} S_{n_A n_A}(\boldsymbol{f})|H_2(\boldsymbol{f})|^2. \quad (11)$$

The averaged cross-power spectral density between the exact displacement signal and the reconstructed displacement signal is given by

$$\bar{S}_{d'd}(\boldsymbol{f}) = \frac{1}{T_{d,x} T_{d,y}} S_{dd}(\boldsymbol{f}) H_1(\boldsymbol{f}) H_2(\boldsymbol{f}). \quad (12)$$

Inserting now Eqs. (11) and (12) into Eq. (10) after some simplifications leads to the desired relation between the power spectral density of the exact displacement signal and the power spectral density of the displacement error signal:

$$S_{e_d e_d}(\boldsymbol{f}) = S_{dd}(\boldsymbol{f}) \left| \frac{1}{T_{d,x} T_{d,y}} H_1(\boldsymbol{f}) H_2(\boldsymbol{f}) - 1 \right|^2 + (S_{n_A n_A}(\boldsymbol{f}) + (T_{d,x} T_{d,y}) S_{n_Q n_Q}(\boldsymbol{f})) \times \left| \frac{1}{T_{d,x} T_{d,y}} H_2(\boldsymbol{f}) \right|^2. \quad (13)$$

The performance of the displacement estimation technique is then assessed by the error variance $\sigma^2_{e_d}$ which can be calculated using Parseval's relation

$$\sigma^2_{e_d} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S_{e_d e_d}(f_x, f_y)\, \mathrm{d}f_y\, \mathrm{d}f_x. \quad (14)$$

2.6. Estimation and reconstruction filter frequency responses

To examine the different 2D transformation models, the frequency responses of the reconstruction filter for nearest neighbour interpolation, the affine, and the bilinear interpolation, have been calculated (see Appendix A):

$$H_{2,\mathrm{nearest}}(f_x, f_y) = T_{d,x} T_{d,y} \operatorname{sinc}(\pi T_{d,x} f_x) \operatorname{sinc}(\pi T_{d,y} f_y), \quad (15)$$

$$H_{2,\mathrm{bilin}}(f_x, f_y) = T_{d,x} T_{d,y} \operatorname{sinc}^2(\pi T_{d,x} f_x) \operatorname{sinc}^2(\pi T_{d,y} f_y), \quad (16)$$

$$H_{2,\mathrm{affine}}(f_x, f_y) = T_{d,x} T_{d,y} \operatorname{sinc}(\pi T_{d,x} f_x) \operatorname{sinc}(\pi T_{d,y} f_y) \times \operatorname{sinc}(\pi (T_{d,x} f_x + T_{d,y} f_y)), \quad (17)$$

with $\operatorname{sinc}(\cdot) = \sin(\cdot)/(\cdot)$. In quantitative evaluations the performance of these filters will be compared to the ideal low pass and the optimum Wiener filter:

$$H_{2,\mathrm{ilp}}(\boldsymbol{f}) = \begin{cases} T_{d,x} T_{d,y} & \text{for } |f_x| \leqslant \dfrac{1}{2T_{d,x}}, \quad |f_y| \leqslant \dfrac{1}{2T_{d,y}}, \\ 0 & \text{else}, \end{cases} \quad (18)$$

$$H_{2,\mathrm{opt}}(f) = \frac{S_{dd}(f)|H_1(f)|^2(T_{d,x}T_{d,y})}{S_{dd}(f)|H_1(f)|^2 + (T_{d,x}T_{d,y})S_{n_on_o}(f) + S_{n_sn_s}(f)}, \quad (19)$$

where Eq. (19) is the well-known formula for the Wiener filter regarding $S_{n_on_o}(f)(T_{d,x}T_{d,y}) + S_{n_sn_s}(f)$ as the noise term.

The frequency responses of the optimum corresponding estimation filter have been calculated such that the squared difference between the exact displacement signal and the reconstructed displacement signal is minimized (see Appendix B):

$$H_{1,\mathrm{nearest}}(f_x,f_y) = \frac{1}{T_{d,x}T_{d,y}}H_{2,\mathrm{nearest}}(f_x,f_y), \quad (20)$$

$$H_{1,\mathrm{bilin}}(f_x,f_y) = \frac{9}{T_{d,x}T_{d,y}}\frac{H_{2,\mathrm{bilin}}(f_x,f_y)}{(a+1)(b+1)}, \quad (21)$$

$$H_{1,\mathrm{affine}}(f_x,f_y) = \frac{6}{T_{d,x}T_{d,y}}\frac{H_{2,\mathrm{affine}}(f_x,f_y)}{a+b+c}, \quad (22)$$

with

$$a = 2\cos^2(\pi T_{d,x}f_x), \quad b = 2\cos^2(\pi T_{d,y}f_y),$$

$$c = 2\cos^2(\pi(T_{d,x}f_x + T_{d,y}f_y)).$$

In the case that $H_2(f)$ is the ideal low pass filter, the estimation filter $H_1(f)$ is also the ideal low pass filter:

$$H_{1,\mathrm{ilp}}(f) = \begin{cases} 1 & \text{for } |f_x| \leqslant \frac{1}{2T_{d,x}},\ |f_y| \leqslant \frac{1}{2T_{d,y}}, \\ 0 & \text{else.} \end{cases} \quad (23)$$

3. Extension of the analytical description by consideration of displacement coding

In the previous section an analytical description for displacement estimation has been presented, and this gives an insight into the sources of errors, namely errors due to a restricted displacement vector amplitude resolution, errors due to a restricted spatial resolution of the displacement vector field, and errors due to non-optimum displacement estimation and reconstruction filter. As a quality measure for the reconstructed displacement vector field the displacement error variance has been used.

In this section the analytical description is extended to allow an objective assessment of the efficiency of the displacement estimation techniques in a video coding system by their rate-distortion functions. The rate-distortion function of a displacement estimation technique describes the relationship between the encoding bit-rate required for transmission of an estimated sparse displacement vector field and the corresponding theoretical lower bound of its displacement error.

The transmission rate needed to code an estimated sparse displacement vector field is affected by its spatial resolution and by its amplitude resolution. Assuming that the signal to be coded consists of real-valued samples (amplitude-continous source), rate-distortion theory may be used to determine a rate-distortion function, which relates a predefined transmission rate for this signal (in bits/sample) to the corresponding minimum variance of the coding noise. The coding noise shall include all errors introduced during coding. It is related to the sampled signal. For an objective assessment of the spatial resolution of estimated sparse displacement vector fields, it is thus necessary to combine rate-distortion theory with the analytical description presented in the previous section. The relation between transmission rate and displacement reconstruction error variance then serves for an objective comparison of different spatial resolutions.

The coding process can be described by passing the signal to be transmitted, e.g. the sampled displacement signal, through a filter with the coder transfer function $G(f)$ and the addition of estimated coding noise $S_{n_cn_c}(f)$, assumed to be independent of the signal.

Similar to Fig. 1 we now obtain a block diagram which describes the sampling, coding, and reconstruction of a space- and amplitude-continuous displacement vector field (Fig. 2). It is straightforward to determine the power spectral density of the displacement error signal in relation to the exact displacement signal, the frequency responses of the estimation and reconstruction filter, the coder transfer function, and the coding noise signal:

$$S_{e_de_d}(f) = S_{dd}(f)\left|\frac{1}{T_{d,x}T_{d,y}}H_1(f)G(f)H_2(f) - 1\right|^2 + S_{n_sn_s}(f)\left|\frac{1}{T_{d,x}T_{d,y}}G(f)H_2(f)\right|^2 + (T_{d,x}T_{d,y})S_{n_cn_c}(f)\left|\frac{1}{T_{d,x}T_{d,y}}H_2(f)\right|^2. \quad (24)$$

The overall displacement error variance $\sigma^2_{e_d}$ can then be calculated according to Eq. (14).

The coder transfer function and the corresponding coding noise depend on the coder source model. In the following two source models, memoryless Gaussian source and Gaussian source with memory, will be considered.

3.1. Encoding of displacement vector fields assuming a memoryless Gaussian source

For the case of encoding displacement vector fields with the model of a zero-mean Gaussian memoryless source, the minimum transmission rate for a predefined distortion Θ is given by [9]

$$R_{G,\mathrm{mle}}(\Theta) = 2\cdot\frac{1}{2}\max\left[0, \log_2\frac{\sigma^2_{d_s}}{\Theta}\right] \text{ bit per coded vector,} \quad (25)$$

where the subscript G,mle denotes the encoding of a memoryless Gausssian source. The factor 2 in Eq. (25) considers the fact that two components of the displacement vector field with equal variance

$$\sigma^2_{d_s} = \frac{1}{f_{d,x}f_{d,y}}\int_{-f_{d,x}/2}^{f_{d,x}/2}\int_{-f_{d,y}/2}^{f_{d,y}/2} S_{d_sd_s}(f_x,f_y)\,df_y\,df_x \quad (26)$$

have to be coded where $f_{d,x}$ and $f_{d,y}$ denote the horizontal and vertical sampling frequencies of the exact displacement signal, i.e. the reciprocal of the sampling periods $T_{d,x}$ and $T_{d,y}$. It is obvious from Eq. (25) that no information needs to be transmitted when the variance of the distortion Θ is larger than the variance of the sampled displacement signal $\sigma^2_{d_s}$.

The minimum transmission rate is achieved by scaling the sampled displacement signal by a filter with a transfer function

$$G_{\mathrm{mle}}(f) = \max\left[0, 1 - \frac{\Theta}{\sigma^2_{d_s}}\right], \quad (27)$$

followed by an addition of white Gaussian noise with a power spectral density

$$S_{\mathrm{mle},n_cn_c}(f) = \Theta\max\left[0, 1 - \frac{\Theta}{\sigma^2_{d_s}}\right]. \quad (28)$$

By variation of the parameter Θ and exploitation of Eqs. (14) and (25) we can construct the desired rate-distortion curve $R_{G,\mathrm{mle}}(\sigma^2_{e_d})$.

An interesting interpretation of (25) is that the non-zero rate is given by the difference in entropies of the source and the noise, which are two zero-mean memoryless Gaussian random variables with variances of $\sigma^2_{d_s}$ and Θ. As the Gaussian source, among others with equal variance, has the maximum entropy, (25) gives an upper bound for non-Gaussian sources.

3.2. Encoding of displacement vector fields assuming a Gaussian source with memory

Sources with memory permit greater data compression than memoryless sources. The redundancy may be removed either in the spatial domain or in the frequency domain. In order to be independent

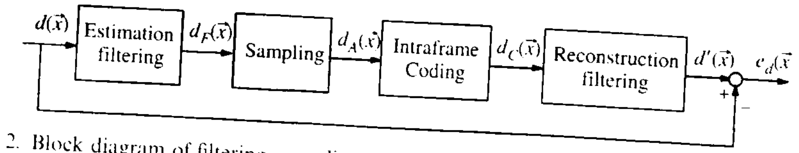


Fig. 2. Block diagram of filtering, sampling, coding and reconstruction of a displacement signal.

of the specific implementation of the encoder used for the displacement vector field encoding, assume that the displacement vector field to be coded is passed through the 'optimum forward channel'. This channel describes the transformation between coder input and decoder output for the lowest transmission rate with respect to a predefined distortion. For the case of a zero-mean Gaussian source with memory, it consists of a nonideal band-limiting filter with a transfer function [9]

$$G_{\mathrm{opt}}(f) = \max\left[0, 1 - \frac{\Theta}{S_{d_\mathrm{A}d_\mathrm{A}}(f)}\right], \tag{29}$$

followed by an addition of a band-limited, non-white Gaussian noise with a power spectral density

$$S_{\mathrm{opt},n_\mathrm{c}n_\mathrm{c}}(f) = \Theta \max\left[0, 1 - \frac{\Theta}{S_{d_\mathrm{A}d_\mathrm{A}}(f)}\right]. \tag{30}$$

Now for the assumption that the two components of the displacement vector field have a Gaussian joint probability density function (p.d.f.) and equal variance, the minimum transmission rate for the two components with distortion Θ is given by [9]

$$R_{\mathrm{G.opt}}(\Theta) = 2\frac{1}{f_{d.x}f_{d.y}}\int_{-f_{d.x}/2}^{f_{d.x}/2}\int_{-f_{d.x}/2}^{f_{d.x}/2} \frac{1}{2}\max\left[0, \log_2\frac{S_{d_\mathrm{A}d_\mathrm{A}}(f_x f_y)}{\Theta}\right]\mathrm{d}f_y\,\mathrm{d}f_x$$

$$\text{bit per coded vector.} \tag{31}$$

It is obvious now, that only frequency-bands with power spectral density above Θ are coded. By varying the parameter Θ we can now analyse the rate-distortion functions of the given idealised coders for various sampling ratios and displacement reconstruction filters.

Again, (31) gives an upper bound for non-Gaussian sources with equal power spectral density.

4. Model assumptions and parameters

In the previous sections an analytical description of displacement estimation and coding has been presented.

For a quantitative evaluation of (14) the exact displacement signal and the quantisation noise signal must be described by parametric models which, in the statistic mean, represent their power spectral densities in image sequences.

4.1. Model assumptions

Here an isotropic autocorrelation function has been chosen to describe the exact displacement signal:

$$R_{dd}(\Delta x, \Delta y) = \sigma_d^2\, \mathrm{e}^{-\alpha\sqrt{\Delta x^2 + \Delta y^2}}. \tag{32}$$

This model is characterized by the model parameter σ_d^2 and α. The power spectral density $S_{dd}(f)$ can be calculated from (32) by Fourier transformation of $R_{dd}(\Delta x, \Delta y)$ and band-limitation to the half of the horizontal and vertical sampling frequencies of the image signal $f_{s.x}$ and $f_{s.y}$:

$$S_{dd}(f_x, f_y) = \begin{cases} \dfrac{f_0\sigma_d^2}{2\pi(f_0^2 + f_x^2 + f_y^2)^{3/2}}, & \text{for } |f_x| \le f_{s.x}/2,\ |f_y| \le f_{s.y}/2, \\ 0, & \text{else}, \end{cases} \tag{33}$$

with

$$f_0 = \alpha/2\pi. \tag{34}$$

The relation $f_{s.x}/f_{s.y}$ between the horizontal and vertical sampling frequency results from the image geometry, i.e. the relation H/B between the height and width of the image and the number of lines NLIN and the number of pels/line NPEL of the sampled image

$$\frac{f_{s.x}}{f_{s.y}} = \frac{H/\mathrm{NLIN}}{B/\mathrm{NPEL}}. \tag{35}$$

The horizontal and vertical sampling frequencies of the displacement signal then follow:

$$f_{d.x} = Gf_{s.x}, f_{d.y} = Gf_{s.y}, \tag{36}$$

where G determines the spatial resolution of the displacement field.

For an analytical description of the probability density function (p.d.f.) of the displacement signal a generalized Gaussian p.d.f. [12] has been chosen:

$$p_d(d) = \frac{v\alpha(v)}{2\sigma_d\Gamma(1/v)}\mathrm{e}^{-(\alpha(v)|d|/\sigma_d)^v}, \tag{37}$$

with

$$\alpha(v) = \sqrt{\frac{\Gamma(3/v)}{\Gamma(1/v)}}, \quad \Gamma\text{: Gamma function.}$$

The p.d.f. $p_d(d)$ is characterised by the standard deviation σ_d and the so called shape parameter v. For the special cases $v = 1$ and $v = 2$ the generalized Gaussian becomes a Laplacian or a Gaussian p.d.f., respectively.

4.2. Model parameters

For a quantitative evaluation the test sequences 'Claire', 'Miss America', 'Alexis' and 'Trevor White' in CIF resolution with a frame rate of 10 Hz for 'Claire', 'Miss America' and 'Alexis' and 7.5 Hz for 'Trevor White' have been chosen. Pelwise displacement vector fields have been measured from these sequences using a hierarchical block matching technique [1] with 1/8 pel displacement amplitude resolution. The x-components of the resulting displacement vector fields served for the determination of the power spectral density of the exact displacement signal and the probability density function of the displacment vectors.

The power spectral density $S_{dd}(f)$ has been measured using the Welsh's method [17] for all images of the four test sequences. Furthermore the variance σ_d^2 and the correlation coefficient ϱ_h of horizontally neighboured displacement vectors of the pelwise displacement vector fields have been measured. With

$$\varrho_\mathrm{h} = R_{dd}(f_{s.x}^{-1}, 0)/\sigma_d^2 = \mathrm{e}^{-\alpha/f_{s.x}}, \tag{38}$$

f_0 has been calculated by

$$f_0 = \frac{\alpha}{2\pi} = -\frac{f_{s.x}}{2\pi}\ln\varrho_\mathrm{h}. \tag{39}$$

Based on these measurements f_0 has been chosen to fit the model power spectral density to the measured power spectral density for all frequencies in the base-band on one side and to fit the measured and model correlation coefficients on the other side.

Table 1
Displacement quantisation error variance for various displacement amplitude resolutions for $v = 0.3$ (see Eq. (37))

Displacement amplitude resolution	1	1/2	1/4	1/8
Displacement quantisation error variance	0.041	0.013	0.004	0.0011

In (40) the chosen model parameters are given:

$$\begin{aligned} &H/B = 3/4, && \sigma_d^2 = 0.94, \\ &\mathrm{NLIN/NPEL} = 288/352, && f_0 = 0.018\ \mathrm{MHz}, \\ &f_{s.x} = 6.75\ \mathrm{MHz}, && \varrho_\mathrm{h} = 0.983. \end{aligned} \tag{40}$$

Using again the x-component of estimated displacement vector fields the probability density function of the displacement signal has been measured. Using the displacement variance σ_d^2 as given in (40) the shape parameter v of the model p.d.f. (37) has been determined by graphical comparison to be $v = 0.3$. Then the quantisation error variances for various quantiser step sizes Δ (amplitude resolution of displacement vector field) have been calculated. They are given in Table 1 for $v = 0.3$.

5. Quantitative evaluations

The analytical description is now used for an examination of displacement estimation and displacement vector field coding.

5.1. Examination of displacement estimation

The first two simulations address the performance of the different reconstruction filters. For this, the frequency response of the implicit estimation filter is set to one: $H_1(f) = 1$. No amplitude quantisation was performed: $S_{n_\mathrm{Q}n_\mathrm{Q}}(f) = 0$. In the first simulation the different reconstruction filters are examined for various spatial resolutions. As can be seen in Fig. 3, the displacement error

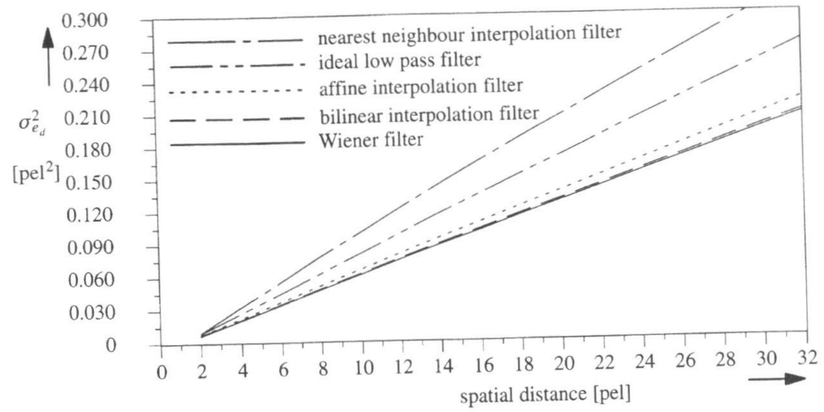


Fig. 3. Calculated displacement error variance for different displacement interpolation filters as a function of the spatial resolution (spatial distance) of the displacement vector field. No estimation filter was used ($H_1(f) = 1$).

variance $\sigma^2_{e_d}$ increases almost linearly with the spatial distance of samples. The lowest error is achieved with the Wiener filter. The affine and bilinear interpolation achieve a performance similar to the Wiener filter. The displacement error variance for nearest neighbour interpolation is almost 2 times the error variance of the affine and bilinear interpolation.

The second simulation addresses the influence of the amplitude resolution of the estimated vector field. The frequency response of the implicit estimation filter is still set to one: $H_1(f) = 1$. A comparison between different amplitude resolutions of the displacement vectors for different spatial resolutions and for bilinear and nearest neighbour interpolation is given in Fig. 4. It can be seen that the displacement error variance decreases with the increase of the amplitude resolution of the displacement vectors. A large gain is reached by using half pel resolution instead of full pel resolution, whereas higher resolution than half pel leads to only marginal benefits. In general, the displacement error variance is more affected by the amplitude resolution in the case of nearest

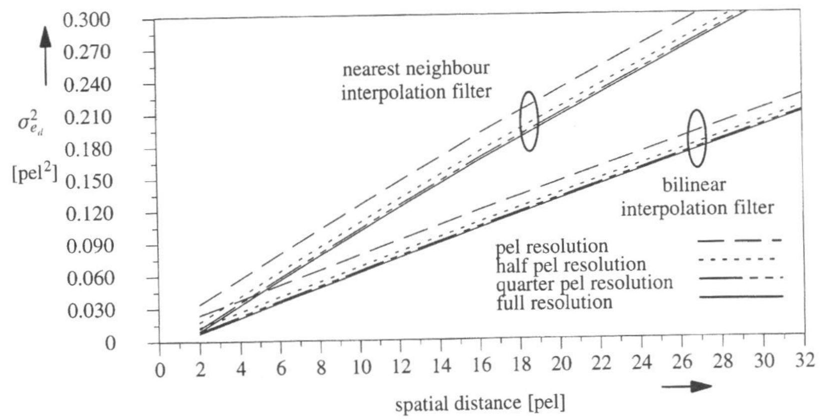


Fig. 4. Calculated displacement error variance for nearest neighbour and bilinear displacement interpolation filters as a function of the spatial resolution (spatial distance) of the displacement vector field. Parameter is the displacement amplitude resolution. No estimation filter was used ($H_1(f) = 1$).

Table 2
Measured and calculated displacement error variances. Averaged values for the image sequences CIF, 10 Hz: 'Alexis', 'Claire', 'Miss America', and CIF, 7.5 Hz: 'Trevor White'. Spatial distance of displacement vectors: 16 × 16 pel

	Nearest neighbour interpolation		Bilinear interpolation	
Displacement error variance	Measured	Calculated	Measured	Calculated
Full pel amplitude resolution	0.184	0.190	0.111	0.119
Half pel amplitude resolution	0.172	0.174	0.107	0.108
Quarter pel amplitude resolution	0.169	0.168	0.105	0.104

neighbour interpolation than in the case of bilinear interpolation.

To confirm these theoretical results, for the third simulation the pelwise given displacement vector fields used for the definition of the model power spectral density were sampled and reconstructed using nearest neighbour and bilinear interpolation. The resulting displacement error variances $\sigma^2_{e_d}$ for these cases were measured. Exemplary results are given in Table 2 for nearest neighbour interpolation and bilinear interpolation with various amplitude resolutions of the displacement vectors and a spatial distance of 16 × 16 pel.

As can be seen the calculated figures coincide well with the measured figures for fine displacement amplitude resolutions. The small differences which can be noticed for full pel amplitude resolution are due to the description of the quantisation. Although the coarse step size has been considered in the quantiser model for the calculation of the quantisation error variance, its influence on the autocorrelation of the displacement error signal and on the cross-correlation between the displacement signal and the displacement error signal has been neglected. The displacement error signal has been assumed to have a white power spectral density (no autocorrelation) and to be additive to the displacement signal (no cross-correlation). This explains the small errors for coarser displacement amplitude resolutions.

The fourth simulation addresses the influence of the implicit estimation filter. A comparison between different amplitude resolutions of the displacement vectors for different spatial resolutions and for bilinear and nearest neighbour interpolation is given in Fig. 5.

A comparison of Fig. 4 with Fig. 5 shows a significant reduction of the displacement error variance

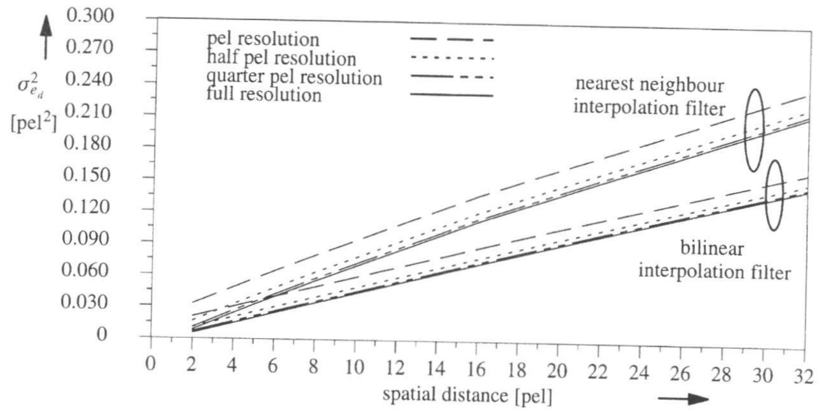


Fig. 5. Calculated displacement error variance for nearest neighbour and bilinear displacement interpolation filters as a function of the spatial resolution (spatial distance) of the displacement vector field. Various displacement amplitude resolutions. Parameter is the displacement amplitude resolution. The estimation filter were chosen corresponding to the interpolation filter.

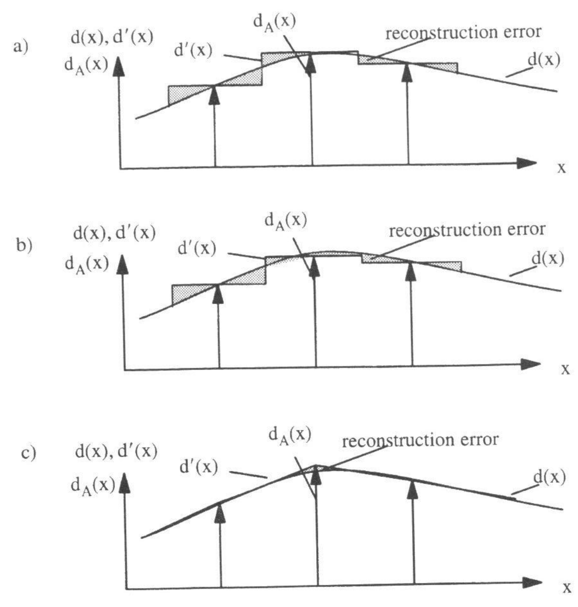


Fig. 6. Illustration of the displacement reconstruction error for: (a) $H_1(f) = 1$ and $H_2(f) = H_{2,\mathrm{nearest}}(f)$, (b) $H_1(f) = H_{1,\mathrm{nearest}}(f)$ and $H_2(f) = H_{2,\mathrm{nearest}}(f)$, (c) $H_1(f) = H_{1,\mathrm{bilin}}(f)$ and $H_2(f) = H_{2,\mathrm{bilin}}(f)$. One-dimensional signals of the continuous displacement signal $d(x)$, the sampled displacement signal $d_A(x)$, and the reconstructed displacement signal $d(x)$ are shown. No displacement amplitude quantisation is assumed.

due to the estimation filter. This fact can be explained by Fig. 6. It shows how the reconstruction error is formed by using different estimation and reconstruction filter. When no estimation filter is used ($H_1(f) = 1$, Fig. 6(a)) the reconstructed displacement signal is exact at the sampling points. At all other points a significant reconstruction error might occur and the displacement error variance might be large (see Fig. 6(a) for the case of $H_2(f) = H_{2,\mathrm{nearest}}(f)$). The estimation filter is calculated in that way that the displacement error variance of the reconstructed displacement signal is minimized. However, as a result of the minimization the reconstructed displacement signal at the sampling points is no longer exact (see Fig. 6(b) for the case of $H_1(f) = H_{1,\mathrm{nearest}}(f)$ and $H_2(f) = H_{2,\mathrm{nearest}}(f)$). It is now an averaged value of the true displacements in the local neighbourhood. Fig. 6(c) shows that the reconstruction error can be significantly reduced by applying the bilinear estimation and reconstruction filter ($H_1(f) = H_{1,\mathrm{bilin}}(f)$ and $H_2(f) = H_{2,\mathrm{bilin}}(f)$). A comparison between Figs. 6(b) and (c) shows how the sampled signals differ when different estimation filters are used. This strengthens the fact that the same motion compensation method as has been used for the displacement estimation should be used for the motion compensated prediction of the image signal.

As one final result we can see from Fig. 5 that when the estimation filter is considered in the analytical description, the displacement error variance for a spatial distance of 16 pel and full pel amplitude resolutions in case of nearest neighbour interpolation is 1.56 times the error variance of bilinear interpolation.

The results given have been obtained as an average for the four test sequences. Of course the absolute figures given depend on the test sequence under consideration, to be more precise, on the kind and the amount of 2D motion within the scene. If the scene contains mainly 2D translational motion and thus the correlation coefficient of horizontally neighboured samples of the exact displacement field are higher than assumed here, the slope of the curves given in Figs. 3–5 will be smaller. However, the relation between the curves will stay the same. If the scene contains a large amount of 2D motion and thus the displacement variance is higher than assumed here, the displacement error variance will also be higher than that given in Fig. 5 for coarse displacement amplitude resolutions (e.g. full pel displacement amplitude resolutions).

5.2. Examination of displacement vector field coding

The next simulations address the influence of the sampling ratio on the transmission rate needed to achieve a predefined displacement error variance. For these simulations the frequency response of the implicit estimation filter is set to one: $H_1(f) = 1$. For the case of the reconstruction filter being the nearest neighbour interpolation filter or the bilinear interpolation filter, the resulting rate-distortion functions for encoding the displacement vector with the assumption of a Gaussian source with memory according to Eq. (31) are depicted in Fig. 7.

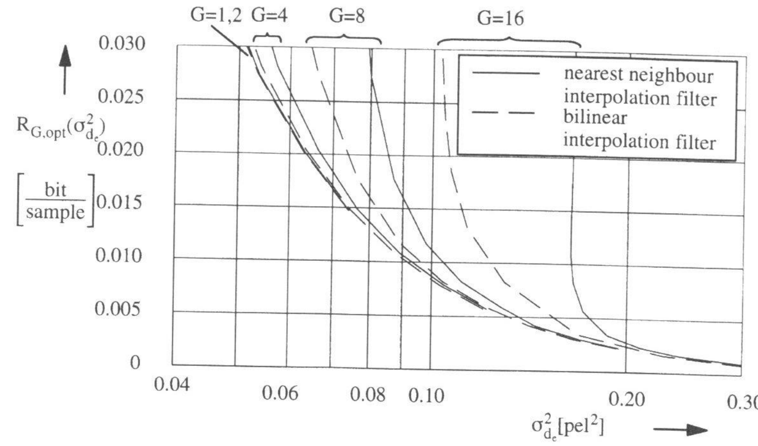


Fig. 7. Rate-distortion functions for nearest neighbour and bilinear displacement interpolation filters. Parameter is the spatial resolution (spatial distance) of the displacement vector field. Displacement vector fields are encoded with the assumption of a Gaussian source with memory (Eq. (31)).

As can be seen the spatial resolution of displacement vectors is a limiting factor for the achievable displacement error variance. Thus, if a low displacement error variance is desired a low spatial distance of displacement vectors has to be chosen. For high displacement error variances the different distortion functions merge. For the quality criterion of displacement error variance it is unimportant whether errors are included due to displacement vector field sampling or due to displacement vector amplitude quantisation. It can be seen further that for each spatial resolution of the displacement vector field, the rate-distortion function for the bilinear interpolation filter is below the corresponding rate-distortion function for the nearest neighbour interpolation filter.

For an optimum coding the spatial resolution of the vector fields must be chosen in such a way that the working point of the encoder lies on the lower bounding curve. This lower bounding curve is the same as the given rate-distortion curve for spatial distance of displacement vectors of one. However, as can be seen in Fig. 7 for a distortion of $\sigma^2_{e_d} \geqslant 0.19$ in the case of nearest neighbour interpolation and $\sigma^2_{e_d} \geqslant 0.12$ in the case of bilinear interpolation, a spatial distance of displacement vectors of 16 is sufficient. Regarding again Table 2 we see that these limits are given by 1 pel displacement vector amplitude resolution. For coding displacement vector fields with less distortion, it seems to be advantageous to increase the spatial resolution instead of the amplitude resolution of the displacement vectors in order to code a displacement vector field at a minimum transmission rate. Of course in a real codec the influence of the displacement error onto the prediction error image has to be taken into account too. One way to find a theoretically optimum bit allocation for the displacement information and the prediction error might be to combine the analytical description presented here with the analytical description of motion compensating prediction presented in [4].

To compare these theoretical results with the transmission rates for displacement vector fields in todays video coding standards, rate-distortion values have been measured. For this purpose again the pelwise given displacement vector fields used for the definition of the model power spectral density were sampled with a spatial distance of 16 and 8 pel and the amplitude has been quantised with amplitude resolutions of 1, 1/2 and 1/4 pel. Distortion values for a spatial distance of 16 pel have already been given in Table 2. To measure the transmission rate, in one simulation the sampled and quantised displacement vector fields have been coded without spatial prediction, whereas in another simulation the spatial displacement vector prediction method of the H.263 [8] video coding

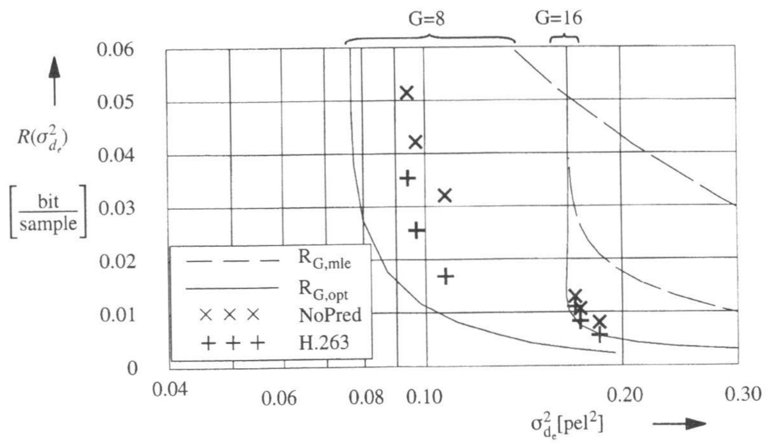


Fig. 8. Measured and calculated rate-distortion functions. The measured rate-distortion values are for displacement amplitude resolutions of 1, 1/2 and 1/4 pel. H.263-like spatial prediction is compared to no spatial prediction (NoPred). The calculated rate-distortion functions are for the assumption of a memoryless Gaussian source ($R_{G,mle}(\sigma^2_{d_e})$, Eq. (25)), and the assumption of a Gaussian source with memory ($R_{G,opt}(\sigma^2_{d_e})$, Eq. (31)). The nearest neighbour interpolation filter was used.

standard has been applied. According to the H.263 video coding standard a displacement vector prediction is obtained by first quantising the displacement vectors to full pel, half pel or quarter pel amplitude resolution and then calculating a median displacement vector of the left, upper and upper left neighbour. To be independent of the implemented Huffman coding tables, which cannot be optimum, the entropy of the error between the displacement vector prediction and the quantised original displacement vectors has been measured. This entropy gives the lower bound for the transmission rate achievable by entropy encoding, e.g. Huffman coding.

For a qualitative assessment of the measured entropy values and corresponding distortion values, the rate-distortion curves for the assumption of a memoryless Gaussian source (Eq. (25)), and a Gaussian source with memory (Eq. (31)) have been calculated. To be conform with the manner how the measured rate-distortion values have been obtained, for these simulations the nearest neighbour interpolation filter, but no implicit estimation filter, was used ($H_1(f) = 1$). A comparison of the rate-distortion curves is given in Fig. 8.

As can be seen in Fig. 8 the measured rate-distortion points given by coding the displacement vector field without spatial prediction are below the rate-distortion curve for the assumption of a memoryless Gaussian source $R_{G,mle}(\sigma^2_{d_e})$. Remember that $R_{G,mle}(\sigma^2_{d_e})$ gives an upper bound for signals with the same power spectral density but non-Gaussian p.d.f., when still a Gaussian p.d.f. of the quantisation error signal is assumed. As we know the p.d.f. of displacement vector fields is non-Gaussian.

By exploiting the spatial correlation of displacement vector fields a reduction of transmission rate is achieved. Again, $R_{G,opt}(\sigma^2_{d_e})$ gives an upper bound for signals with non-Gaussian p.d.f., when still Gaussian p.d.f. of the quantisation error signals is assumed. Consequently, if Gaussian p.d.f. of the quantisation error signals can further be assumed, an optimum exploitation of the spatial correlations of displacement vector fields drops transmission rates to $R_{G,opt}(\sigma^2_{d_e})$ or even below.

6. Conclusions

An analytical description of displacement estimation is presented which allows an objective evaluation and comparison of various displacement estimation techniques. An additional value of the analytical description of displacement estimation is seen in its potential for an in-depth understanding of the different error sources of displacement estimation. Displacement estimation is described analytically by low pass filtering, spatial sampling, quantisation and reconstruction of an exact displacement vector field.

For the purpose of analysis the rate-distortion theory is applied. Distortion is introduced by a limited displacement vector amplitude resolution (quantisation) and a limited spatial resolution of displacement vector fields (sampling). From the view of this general description of displacement vector field coding, the term 'lossless encoding of displacement vector fields' as used frequently in the literature describes the encoding of displacement information which is already distorted. The displacement reconstruction filter applied to transmitted displacement vector fields reduces the distortion, i.e. the displacement error variance.

Comparisons of various displacement reconstruction techniques applied to displacement vector fields with various amplitude and spatial resolutions show that the bilinear interpolation filter performs always better than the nearest neighbour interpolation filter and very close to optimum Wiener filter. The affine interpolation filter performs slightly worse than the bilinear interpolation filter. The theoretical results obtained for various displacement vector amplitude resolutions, various spatial resolutions, and various displacement reconstruction techniques confirm the results achieved by heuristics.

During displacement estimation a local neighbourhood in the image signal is used to determine one displacement vector. Thus, if the picture elements in the local neighbourhood have been displaced differently the estimated displacement vector is not a sample of the displacement signal, but is an averaged value of the displacements in the local neighbourhood. This effect is described by passing the displacement signal through an estimation filter before sampling. For each reconstruction filter there exists a corresponding optimal estimation filter, the transfer function of which is presented. Simulations show that a corresponding pair of estimation and reconstruction filter leads to a significant decrease of the displacement error variance. This means for practical implementations that the same reconstruction filter as has been used for the displacement estimation must be used for the motion compensated prediction of the image signal.

The rate-distortion analysis shows that allowing an infinite displacement vector amplitude resolution, the spatial resolution of the displacement vector field is a limiting factor for the achievable distortion, i.e. the displacement error variance. Thus, if a low displacement error variance is desired a high spatial resolution of displacement vector fields has to be chosen. The rate-distortion analysis confirms that for a full pel displacement vector amplitude resolution a spatial distance of displacement vectors of 16×16 pel is a good choice. For coding displacement vector fields with a lower distortion, the results propose to increase the spatial resolution instead of the amplitude resolution of the displacement vectors in order to code a displacement vector field at a lower transmission rate.

In a video coding system the influence of the displacement error onto the motion compensated prediction error of the luminance and chrominance samples has to be taken into account additionally. In order to find the theoretically optimum bit allocation for the displacement information and the prediction error image the analytical description presented here can be combined with the analytical description of motion compensating prediction presented in [4].

Acknowledgements

The author wishes to thank Prof. Dr.-Ing. H.G. Musmann and Dr.-Ing. Bernd Edler of the Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, Universität Hannover, for many helpful discussions. Grateful acknowledgement is made to Peter Gerken, also of the Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung for his detailed comments on the various drafts of this paper. This research was supported in part by the European RACE project R2110 HAMLET.

Appendix A. Calculation of reconstruction filter frequency responses

The reconstruction filter serves for the spatial interpolation of a sparse displacement vector field. The interpolation techniques under consideration here, are the affine and bilinear interpolation of estimated displacement vectors at node locations of a regular triangular or quadrilateral mesh. The attachment of a single estimated displacement vector to a whole block in traditional block matching is viewed as nearest neighbour interpolation of displacement vectors at node locations of a regular quadrilateral mesh.

To describe the reconstruction of the displacement signal within one mesh element (patch) of a regular quadrilateral mesh a local coordinate system with coordinates X and Y is introduced here. As origin of the local coordinate system, the coordinate of the upper left node of the patch has been chosen. Its coordinates $(x = X_0,\ y = Y_0)$ are then given by the coordinates $(x = x_0,\ y = y_0)$, of a point for which a displacement vector shall be calculated:

$$X_0 = \mathrm{INT}(x_0/M)\,M, \qquad Y_0 = \mathrm{INT}(y_0/N)\,N, \tag{A.1}$$

where $\mathrm{INT}(\cdot)$ denotes the integer operation and M, N the horizontal and vertical distance of nodes of the mesh. The distance of nodes of the mesh corresponds to the horizontal and vertical sampling period of the displacement signal

$$M = T_{d,x}, \qquad N = T_{d,y}. \tag{A.2}$$

The local coordinates of the point under consideration are then given by

$$X = x_0 - X_0, \qquad Y = y_0 - Y_0. \tag{A.3}$$

For reconstruction of $d'(X, Y)$ from the quantised four nearest samples $d_{0,0} = d_Q(0, 0)$, $d_{M,0} = d_Q(M, 0)$, $d_{0,N} = d_Q(0, N)$ and $d_{M,N} = d_Q(M, N)$ of the filtered displacement signal (Fig. 9), by use of nearest neighbour interpolation, bilinear interpolation, or affine interploation, we get

$$d'_{\mathrm{nearest}}(X, Y) = \begin{cases} d_{0,0} & \text{for } 0 \leqslant X < M/2, \quad 0 \leqslant Y < N/2, \\ d_{M,0} & \text{for } M/2 \leqslant X < M, \quad 0 \leqslant Y < N/2, \\ d_{0,N} & \text{for } 0 \leqslant X < M/2, \quad N/2 \leqslant Y < N, \\ d_{M,N} & \text{for } M/2 \leqslant X < M, \quad N/2 \leqslant Y < N, \end{cases} \tag{A.4}$$

$$d'_{\mathrm{bilin}}(X, Y) = \left(1 - \frac{X}{M}\right)\left(1 - \frac{Y}{N}\right) d_{0,0} + \frac{X}{M}\left(1 - \frac{Y}{N}\right) d_{M,0} + \left(1 - \frac{X}{M}\right)\frac{Y}{N} d_{0,N} + \frac{X}{M}\frac{Y}{N} d_{M,N} \quad \text{for } 0 \leqslant X < M, \quad 0 \leqslant Y < N, \tag{A.5}$$

$$d'_{\mathrm{affine}}(X, Y) = \begin{cases} \left(1 - \dfrac{X}{M}\right) d_{0,0} + \left(\dfrac{X}{M} - \dfrac{Y}{N}\right) d_{M,0} + \dfrac{Y}{N} d_{M,N} & \text{for } 0 \leqslant X < M, \quad 0 \leqslant Y < X\dfrac{N}{M}, \\ \left(1 - \dfrac{Y}{N}\right) d_{0,0} + \left(\dfrac{Y}{N} - \dfrac{X}{M}\right) d_{0,N} + \dfrac{X}{M} d_{M,N} & \text{for } 0 \leqslant X < M, \quad X\dfrac{N}{M} \leqslant Y < N. \end{cases} \tag{A.6}$$

Regarding the general description for two dimensional filtering,

$$d'(X, Y) = d_Q(X, Y) * h_2(X, Y) = \iint d_Q(X', Y')\,h_2(X - X', Y - Y')\,\mathrm{d}X'\,\mathrm{d}Y', \tag{A.7}$$

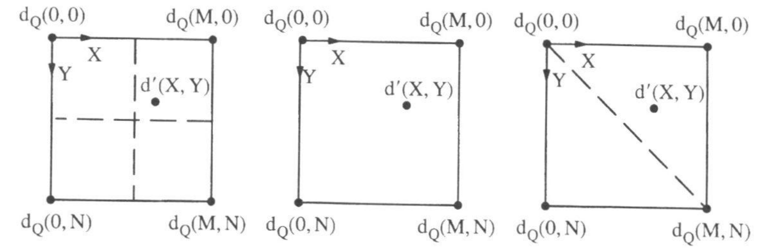


Fig. 9. Local neighbourhood for the reconstruction of a displacement vector using (a) nearest neighbour interpolation, (b) bilinar interpolation and (c) affine interpolation.

the filter impulse responses are obtained by comparing the coefficients

$$h_{2,\mathrm{nearest}}(X, Y) = \begin{cases} 1 & \text{for } 0 \leqslant |X| < M/2, \quad 0 \leqslant |Y| < N/2, \\ 0 & \text{else}, \end{cases} \tag{A.8}$$

$$h_{2,\mathrm{bilin}}(X, Y) = \begin{cases} (1 - X/M)(1 - Y/N) & \text{for } 0 \leqslant X < M, \quad 0 \leqslant Y < N, \\ (1 + X/M)(1 - Y/N) & \text{for } -M \leqslant X < 0, \quad 0 \leqslant Y < N, \\ (1 - X/M)(1 + Y/N) & \text{for } 0 \leqslant X < M, \quad -N \leqslant Y < 0, \\ (1 + X/M)(1 + Y/N) & \text{for } -M \leqslant X < 0, \quad -N \leqslant Y < 0, \\ 0 & \text{else}, \end{cases} \tag{A.9}$$

$$h_{2,\mathrm{affine}}(X, Y) = \begin{cases} 1 - X/M & \text{for } 0 \leqslant X < M, \quad 0 \leqslant Y < X\dfrac{N}{M}, \\ 1 + X/M - Y/N & \text{for } -M \leqslant X < 0, \quad 0 \leqslant Y < X\dfrac{N}{M} + N, \\ 1 + Y/N & \text{for } -M \leqslant X < 0, \quad -N \leqslant Y < X\dfrac{N}{M}, \\ 1 - Y/N & \text{for } 0 \leqslant X < M, \quad X\dfrac{N}{M} \leqslant Y < N, \\ 1 + Y/N - X/M & \text{for } 0 \leqslant X < M, \quad X\dfrac{N}{M} - N \leqslant Y < 0, \\ 1 + X/M & \text{for } -M \leqslant X < 0, \quad X\dfrac{N}{M} \leqslant Y < 0, \\ 0 & \text{else}. \end{cases} \tag{A.10}$$

Now the frequency responses $H_2(f)$ can be calculated by Fourier Transformation of the reconstruction filter impulse responses:

$$H_{2,\mathrm{nearest}}(f_x, f_y) = MN\,\mathrm{sinc}(\pi M f_x)\,\mathrm{sinc}(\pi N f_y), \tag{A.11}$$

$$H_{2,\mathrm{bilin}}(f_x, f_y) = MN\,\mathrm{sinc}^2(\pi M f_x)\,\mathrm{sinc}^2(\pi N f_y), \tag{A.12}$$

$$H_{2,\mathrm{affine}}(f_x, f_y) = MN\,\mathrm{sinc}(\pi M f_x)\,\mathrm{sinc}(\pi N f_y)\,\mathrm{sinc}(\pi(M f_x + N f_y)). \tag{A.13}$$

Eqs. (15)–(17) then follow with (A.2).

Appendix B. Calculation of estimation filter frequency responses

The estimated displacement vector in general is not a sample of the true displacement signal, but is an averaged value of the displacements in the local neighbourhood. This effect has been modelled by low pass filtering the true displacement signal before sampling.

To simplify the calculations it is assumed in the following that displacement estimation can be expressed as a minimisation of the squared difference between the exact displacement signal and a displacement signal, which is implicitly interpolated from the sparse displacement vector field to be estimated. This implicit displacement signal depends on the 2D transformation model of the displacement estimator. The assumption of minimizing the squared difference between the two displacement signals is valid, when during displacement estimation the squared difference between the motion compensated previous frame and the original current frame is evaluated in large measurement windows.

Now we have to determine the frequency response $H_1(f_x, f_y)$ for each frequency $\boldsymbol{f}_0 = (f_{0,x}, f_{0,y})^{\mathrm{T}}$ in such a way, that the displacement error variance according to Eq. (14) with (13) and (4) becomes minimum:

$$\sigma_{e_d}^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Bigg[S_{dd}(f_x, f_y)\left(\frac{1}{T_{d,x}T_{d,y}}H_1(f_x, f_y)H_2(f_x, f_y) - 1\right)^2$$
$$+ \sum_{\substack{n=-\infty \\ (m,n)\neq(0,0)}}^{\infty}\sum_{m=-\infty}^{\infty} S_{dd}\left(f_x + \frac{m}{T_{d,x}}, f_y + \frac{n}{T_{d,y}}\right)\left(\frac{1}{T_{d,x}T_{d,y}}H_1\left(f_x + \frac{m}{T_{d,x}}, f_y + \frac{n}{T_{d,y}}\right)H_2(f_x, f_y)\right)^2$$
$$+ (T_{d,x}T_{d,y})S_{n_Q n_Q}(f_x, f_y)\left(\frac{1}{T_{d,x}T_{d,y}}H_2(f_x, f_y)\right)^2\Bigg]\mathrm{d}f_x\,\mathrm{d}f_y \underset{H_1(f_x, f_y)}{\longrightarrow} \mathrm{MIN}. \tag{B.1}$$

Please remember, that all considered frequency responses are real. By partial derivation of $\sigma_{e_d}^2$ with respect to $H_1(\boldsymbol{f}_0)$ we get, as long as $H_2(\boldsymbol{f})$ does not depend on $H_1(\boldsymbol{f})$, a condition for $H_1(\boldsymbol{f})$:

$$\frac{1}{2}\frac{\partial\sigma_{e_d}^2}{\partial H_1(\boldsymbol{f}_0)} = S_{dd}(\boldsymbol{f}_0)\left(\frac{1}{T_{d,x}T_{d,y}}H_1(\boldsymbol{f}_0)H_2(\boldsymbol{f}_0) - 1\right)\frac{1}{T_{d,x}T_{d,y}}H_2(\boldsymbol{f}_0)$$
$$+ S_{dd}(\boldsymbol{f}_0)H_1(\boldsymbol{f}_0)\sum_{\substack{n=-\infty \\ (m,n)\neq(0,0)}}^{\infty}\sum_{m=-\infty}^{\infty}\left(\frac{1}{T_{d,x}T_{d,y}}H_2\left(f_{0,x} - \frac{m}{T_{d,x}}, f_{0,y} - \frac{n}{T_{d,y}}\right)\right)^2$$
$$\overset{!}{=} 0 \quad \text{for each } \boldsymbol{f}_0 = (f_{0,x}, f_{0,y})^{\mathrm{T}}, \tag{B.2}$$

which we can solve for $H_1(\boldsymbol{f})$:

$$H_1(\boldsymbol{f}) = \frac{H_2(\boldsymbol{f})}{\dfrac{1}{(T_{d,x}T_{d,y})}\displaystyle\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\left(H_2\left(f_x - \frac{m}{T_{d,x}}, f_y - \frac{n}{T_{d,y}}\right)\right)^2}. \tag{B.3}$$

For an evaluation of Eq. (B.3) it is useful to simplify the denominator, because of the infinite summation. The denominator describes normalised shifted copies of $(H_2(f_x, f_y))^2$. It is thus the Fourier transform of a sampled 'signal' given by $h_2(x, y) ** h_2(-x, -y)$, where the operator $(**)$ denotes two-dimensional convolution. We can thus write for $H_1(\boldsymbol{f})$:

$$H_1(\boldsymbol{f}) = \frac{H_2(\boldsymbol{f})}{\displaystyle\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}(h_2(x, y) ** h_2(-x, -y))\delta(x + mT_{d,x}, y + nT_{d,y})\mathrm{e}^{-2\pi mT_{d,x}f_x}\mathrm{e}^{-2\pi nT_{d,y}f_y}}. \tag{B.4}$$

The reconstruction filter under consideration here have short impulse responses and (B.4) can be evaluated easily. The desired estimation filter frequency responses are given in Eqs. (20)–(22).

References

[1] M. Bierling, "Displacement estimation by hierarchical block matching", *Proc. 3rd SPIE Symp. on Visual Comm. Image Process.*, Cambridge, USA, November 1988, pp. 942–951.

[2] F. Dufaux, Multigrid block matching motion estimation for generic video coding, Ph.D. Thesis, École Polytechnique Fédérale de Lausanne, 1994.

[3] C. Fuh and P. Maragos, "Motion displacement estimation using an affine model image matching", *Optical Eng.*, Vol. 30, No. 7, July 1991, pp. 881–887.

[4] B. Girod, "The efficiency of motion-compensating prediction for hybrid coding of video sequences", *J. Select. Areas Comm.*, Vol. SAC-5, August 1987, pp. 1140–1154.

[5] S. Gupta and A. Gersho, "On fractional motion estimation", *Proc. SPIE Visual Comm. Image Process. '93*, Vol. 2094-I, Cambridge, MA, November 1993, pp. 408–419.

[6] ISO/IEC, Description of MPEG-4, Doc. ISO/IEC JTC1/SC29/WG11 N1410, October 1996.

[7] ISO/IEC, Description of Core Experiments on Coding Efficiency in MPEG-4 Video, Doc. ISO/IEC JTC1/SC29/WG11 N1385, October 1996.

[8] ITU-T Recommendation H.263, Video Coding for Low Bitrate Communication, December 1995.

[9] N.S. Jayant and P. Noll, "*Digital Coding of Waveforms*", Prentice-Hall, Englewood Cliffs, NJ, 1984.

[10] J.W. Kim and S.U. Lee, "On the hierarchical variable block size motion estimation technique for motion sequence coding", *Proc. SPIE Visual Comm. Image Process. '93*, Vol. 2094-I, Cambridge, MA, November 1993, pp. 372–383.

[11] H. Li and R. Forchheimer, "A new motion compensation technique for video compression", *Proc. IEEE Internat. Conf. Acoust. Speech Signal Process.*, Minneapolis, MN, May 1993, pp. 441–444.

[12] F. Müller,"On the motion compensated prediction error using true motion fields", *Proc. IEEE Internat. Conf. on Image Process. (ICIP 94)*, Vol III, Austin, Texas, November 1994, pp. 781–785.

[13] Y. Nakaya and Harashima, "An iterative motion estimation using triangular patches for motion compensation", *Proc. SPIE Visual Comm. Image Process. 91*, Vol. 1605, 1991, pp. 546–556.

[14] J. Nieweglowski and P. Haavisto, A method for displacement compensated video prediction, *Proc. Internat. Workshop on HDTV '93*, Ottawa, Canada, October 1993.

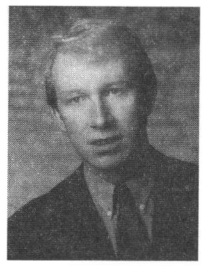
[15] J. Nieweglowski and P. Haavisto, Motion vector field reconstruction for predictive video sequence coding, *Proc. Internat. Workshop on Coding Techniques for Very Low Bit-rate Video*, Colchester, UK, April 1994.

[16] C.A. Papadopoulos, The use of geometric transformations for motion compensation in video data compression, Ph.D. Thesis, University of London, 1994.

[17] A. Papoulis, *Signal Analysis*, McGraw-Hill, New York, 1986.

[18] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 1991.

[19] Y. Wang and O. Lee, "Active mesh – a video representation scheme for feature seeking and tracking", *Proc. SPIE Visual Comm. Image Process. '93*, Vol. 2094, Cambridge, MA, November 1993, pp. 1558–1569.



Ralf Buschmann was born in Bielfeld, Germany, in 1963. He received the Dipl.-Ing. degree in electrical engineering from the University of Hannover, Germany, in 1989. Since 1989, he has been with the Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung at the University of Hannover, where he has been involved in the RACE research projects R2052 "MONA LISA", and R2110 "HAMLET". His current research interest is in displacement estimation for video coding.

General. Prospective authors are encouraged to submit manuscripts within the scope of the Journal. To qualify for publication, papers must be previously unpublished and not be under consideration for publication elsewhere. All material should be sent in quadruplicate (original plus three copies) to the Editor-in-Chief. Contributors are reminded that once their contribution has been accepted for publication, all further correspondence should not be sent to the Editor, but directly to the publishers (Editorial Department, Elsevier Science B.V., P.O. Box 1991, 1000 BZ Amsterdam, The Netherlands).
All manuscripts will be assessed by at least two (anonymous) referees.
Upon acceptance of an article, the author(s) will be asked to transfer copyright of the article to the publisher. This transfer will ensure the widest possible dissemination of information.
Accepted languages are English (preferred), French and German. The text of the paper should be preceded by abstracts of no more than 200 words in English. Abstracts should contain the substance of the methods and results of the paper. Page proofs will be sent to the principal author with an offprint order form. Fifty offprints of each article can be ordered free of charge. Costs arising from alterations in proof, other than of printer's errors, will be charged to the authors. All pages should be numbered. The first page should include the article title and the author's name and affiliation, as well as a name and mailing address to be used for correspondence and transmission of proofs. The second page should include a list of unusual symbols used in the article and the number of pages, tables and figures. It should also contain the keywords in English.
Figures. All illustrations are to be considered as figures, and each should be numbered in sequence with Arabic numerals. The drawings of the figures must be originals, drawn in black india ink and carefully lettered, or printed on a high-quality laser printer. Each figure should have a caption and these should be listed on a separate sheet. Care should be taken that lettering on the original is large enough to be legible after reduction. Each figure should be identified. The approximate place of a figure in the text should be indicated in the margin. In case the author wishes one or more figures to be printed in colour, the *extra* costs arising from such printing will be charged to the author. In this case 200 offprints may be ordered free of charge. More details are available from the Publisher.
Tables. Tables should be typed on separate sheets. Each table should have a number and a title. The approximate places for their insertion in the text should be indicated in the margin.
Footnotes in text. Footnotes in the text should be identified by superscript numbers and listed consecutively on a separate page.
References. References must be *in alphabetical order* in the style shown below:

Book	[1]	A.V. Oppenheim et al., *Digital Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, 1975, Chapter 10, pp. 491–499.
Journal	[2]	F.J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform", *Proc. IEEE*, Vol. 66, No. 1, January 1978, pp. 53–83.
Conference Proceedings	[3]	D. Coulon and D. Kayser, "A supervised-learning technique to identify short natural language sentence", *Proc. 3rd Internat. Joint Conf. on Pattern Recognition*, Coronado, CA, 8–11 November 1976, pp. 85–89.
Contributed Volume	[4]	E.F. Moore, "The firing squad synchronization problem", in: E.F. Moore, ed., *Sequential Machines, Selected Papers*, Addison-Wesley, Reading, MA, 1964, pp. 213–214.

Fast Communications. Papers for the Fast Communications section should be submitted by electronic mail to Prof. Murat Kunt, Laboratoire de Traitement des Signaux, Département d'Electricité, EPFL, Ecublens, CH-1015 Lausanne, Switzerland, Tel.: (4121) 693 26 26, Fax: (4121) 693 46 60, E-mail: fastcom@ltssun17.epfl.ch. Papers should be a maximum of 2,500 words in length (approximately 6 printed journal pages for Signal Processing: Image Communication). Submissions will be subject to the same editorial selection criteria as regular papers. Reviews will be dispatched electronically and decisions will be binary (yes/no) to avoid publication delays. Please ensure your complete postal and e-mail address are indicated on the title page. As no page proofs will be sent to the authors, the presentation should be very clear. For Fast Communications, the figures should be provided in Encapsulated Postscript (eps) format. To ensure fast publication, the manuscript should be written in LaTeX using the document styles of Elsevier Science B.V. Move all files needed (TeX source, eps-files, style and bibliography files) into one directory. Remove all compilation files (*. log, *. lof, *. dvi, *. aux, . . .). Rename the main TeX source into *review.tex* and archive (tar), compress and unencode this directory, and e-mail the unencoded file to: fastcom@ltssun17.epfl.ch. Authors who comply with the above conditions will have their Fast Communication published on the EEE-Alert Server within three weeks of acceptance.

LaTeX files of papers that have been accepted for publication may be sent to the Publisher by e-mail or on a diskette (3.5″ or 5.25″ MS-DOS). If the file is suitable, proofs will be produced without rekeying the text. The article should be encoded in ESP-LaTeX, standard LaTeX, or AMS-LaTeX (in document style 'article'). The ESP-LaTeX package, together with instructions on how to prepare a file, is available from the Publisher. It can also be obtained through the Elsevier WWW home page (http://www.elsevier.nl) or using anonymous FTP from the Comprehensive TeX Archive Network (CTAN). The host-names are: ftp.dante.de, ftp.tex.ac.uk, ftp.shsu.edu; the directory is /tex-archive/macros/latex/contrib/supported/elsevier. *No changes* from the accepted version are permissible, without the explicit approval by the Editors. The Publisher reserves the right to decide whether to use the author's file or not. If the file is sent by e-mail, the name of the journal, *Signal Processing: Image Communication*, should be mentioned in the subject field of the message to identify the paper. Authors should include an ASCII table (available from the Publisher) in their files to enable the detection of transmission errors. The files should be mailed to: Ineke Kolen, Elsevier Science B.V., P.O. Box 103, 1000 AC Amsterdam, The Netherlands. Fax: +31 20 4852829. E-mail: c.kolen@elsevier.nl.

For the purpose of further correspondence the manuscript should end with a complete mailing address, preferably including e-mail address, of at least one of the authors.