Spatially Adaptive Subsampling of Image Sequences

Ricardo A. F. Belfor, *Member, IEEE*, Marc P. A. Hesp,
Reginald L. Lagendijk, *Member, IEEE*, and Jan Biemond, *Fellow, IEEE*

***Abstract*— In a spatially adaptive subsampling scheme, the subsampling lattice is adapted to the local spatial frequency content of an image sequence. In this paper, we use rate-distortion theory to show that spatially adaptive subsampling gives a better performance than subsampling with a fixed sampling lattice. A new algorithm that optimally assigns sampling lattices to different parts of the image is presented. The proposed spatially adaptive subsampling method can be applied within a motion-compensated coding scheme as well. Experiments show an increased performance over fixed lattice subsampling.**

I. Introduction

Subsampling is a basic data compression method in image coding. By discarding a part of the pixels, the image can be transmitted more efficiently. An application of fixed lattice subsampling is the coding of the color information in image sequences [1]. Because of the oversampling of the color information in current video standards, substantial data compression can be obtained. Fixed subsampling, however, discards a part of the spectrum without any consideration as to the actual content of the image. Such an approach causes an unacceptable loss of resolution when applied to luminance information. Therefore, an extension of basic subsampling should be made in order to account for the nonstationary nature of image sequences.

In a spatially adaptive subsampling scheme, the image is subdivided into square blocks, and each block is represented by a specific spatial sampling lattice. In detailed regions, a dense sampling lattice is used, and in regions with little detail, a sampling lattice with only a few pixels is used. The choice of which lattice to use is determined by a rate and quality controller. In [2], the time axis transform (TAT) system that optimally assigns a sampling lattice to each block was described. This algorithm allows only for two different sampling lattices. A nonoptimal solution for the assignment problem in a system with three different sampling lattices was presented in [3]. This method was not optimal because no attempt was made to search for the global minimum distortion, and the algorithm stops in what may be a local minimum. A variation of this algorithm was presented in [4]. The number of different sampling lattices can be extended by using an hierarchical approach [5]. However, this introduces the problem of distributing the bit rate over the different levels in the hierarchy. In this paper, a new algorithm that assigns different spatial sampling lattices to the blocks is described based on rate-distortion theory. This theory has been used successfully in the past for solving the problem of optimally dividing bits among different channels [6]. The advantage of the algorithm is that the set of possible lattices is not limited and that, under certain conditions, optimality can be guaranteed.

Manuscript received April 1, 1993; January 24, 1994. This work was supported by NATO grant 5-2-05/CRG 900834. The associate editor coordinating the review of this paper and approving it for publication was Prof. Dr.-Ing. Bernd Girod.

The authors are with the Department of Electrical Engineering, Information Theory Group, Delft University of Technology, Delft, The Netherlands.

IEEE Log Number 9402257.

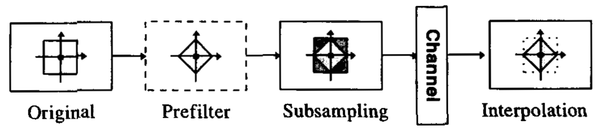


Fig. 1. Basic subsampling scheme. The figures in the boxes illustrate the actions in the spectral domain.

A further compression can be achieved by exploiting the temporal correlation in an image sequence. In a subsampling scheme, this is normally done by using sub-Nyquist subsampling techniques [7], [8]. The approach taken in our scheme is to use motion-compensated prediction as used in hybrid coders. If the prediction error is small, no additional information has to be transmitted. Thus, the spatially adaptive subsampling analogy is extended by decreasing the temporal sampling rate if the temporal activity is low.

In Sections II and III some theoretical background is presented about the basics of subsampling and spatially adaptive subsampling. The purpose of the theory is to provide a mathematical framework for analyzing the performance of adaptive subsampling system. We use the rate-distortion theory to prove that spatially adaptive subsampling works better than fixed lattice subsampling. The practical implementation of a spatially adaptive coding system is described in Sections III and IV, and is finally evaluated in Section V.

II. Fixed Lattice Subsampling

Subsampling can be defined as representing an image sequence on a new sampling lattice with a lower sampling density than the original lattice. A simple subsampling scheme is shown in Fig. 1. To prevent aliasing caused by subsampling, the spatial frequencies of the image sequence should be confined to a unity cell [9]. This is a region in the frequency domain associated with a given sampling lattice and defined in such a way that by tiling this region, a complete coverage of the frequency plane can be obtained without any overlap. This can be seen as an extension of the Nyquist sampling theorem to multiple dimensions on an arbitrary sampling lattice. A prefilter can be used to confine the original image spectrum to

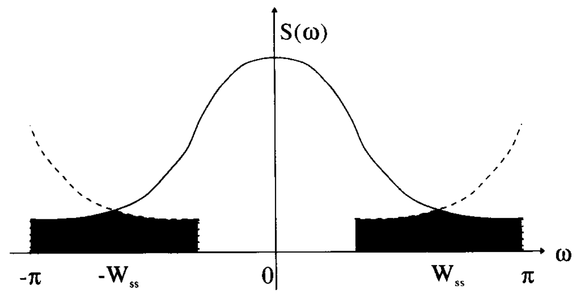


Fig. 2. Power density function. The shaded areas are the components that contribute to the distortion. The dark shade represents the aliasing component and the light shade the loss of resolution.

the unity cell. If the original spectrum is already confined to the unity cell, prefiltering is optional.

If the new sampling lattice is a subset of the original lattice, the actual subsampling can be implemented by simply discarding those pixels not present in the new lattice. If this is not the case, an intermediate sampling structure that bears a relation to both the original and the new lattice must be used [10]. In a subsampling data compression system, the remaining pixels after subsampling are transmitted or forwarded to subsequent coding or processing stages.

At the receiver, the image sequence has to be reconstructed to the original sampling lattice. This is done with an interpolation filter. This filter has to be designed in such a way that the replicas introduced by the subsampling process are cancelled out. Further, it should not remove the frequency components within the unity cell because this would cause a loss of resolution.

We investigate several coding aspects of fixed lattice subsampling from a rate-distortion point of view [11]. Properties derived here are used in the next section when discussing spatially adaptive subsampling. The rate-distortion function $R(D)$ provides a lower bound for the bit rate R necessary to transmit a source with an average distortion D. We derive the rate-distortion function for subsampling as a data reduction method to show the conditions under which subsampling is appropriate and to study the source of the different errors.

We assume a band-limited spatially discrete source that is PCM encoded and has a power spectral density function $S(\omega)$. An example of such a function is shown in Fig. 2. According to Parseval's theorem, the variance σ^2 of this source is

$$\sigma^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} S(\omega)d\omega. \tag{1}$$

In a fixed lattice subsampling scheme, the spectrum of the input source is low-pass filtered and then subsampled according to the Nyquist theorem. If the rate required to transmit the original source using PCM coding is R_0, then the new rate after subsampling is

$$R = \frac{W_{ss}}{\pi} \cdot R_0(\text{bits/pixel}) \tag{2}$$

where W_{ss} is the bandwidth after prefiltering. Hence, the new bit rate is reduced proportionally with the bandwidth reduction.

If we assume that prior to subsampling a prefilter $H(\omega)$ is used with a cut-off frequency at W_{ss}, then the corresponding mean square error distortion function is given by

$$D = \frac{1}{\pi}\int_0^{\pi}(1 - |H(\omega)|^2)S(\omega)d\omega + \frac{1}{\pi}\int_{W_{ss}}^{\pi} |H(\omega)|^2 S(\omega)d\omega \tag{3}$$

where we have taken into account the symmetry of $S(\omega)$ and $H(\omega)$ around the origin. The first part of the equation represents the distortion introduced by the prefilter, and the second part is the aliasing error caused by an imperfect prefilter. The two areas that contribute to the distortion are the shaded regions in Fig. 2. If an ideal prefilter is used with transfer function

$$H(\omega) = \begin{cases} 1, & |\omega| < W_{ss} \\ 0, & W_{ss} < |\omega| < \pi \end{cases} \tag{4}$$

then the aliasing component is zero, and (3) reduces to

$$D = \frac{1}{\pi}\int_{W_{ss}}^{\pi} S(\omega)d\omega. \tag{5}$$

Substituting (2) in (5) and using (1) gives the distortion-rate function $D(R)$:

$$D(R) = \sigma^2 - \frac{1}{\pi}\int_0^{\frac{R}{R_0}\pi} S(\omega)d\omega. \tag{6}$$

Let us consider the properties of this distortion-rate function. The first derivative of $D(R)$ is given by

$$\frac{\partial D}{\partial R} = -\frac{1}{R_0}S\left(\frac{R}{R_0}\pi\right). \tag{7}$$

The first derivative of the rate-distortion function is *always* monotonically decreasing because $S(\omega)$ is always greater or equal to zero. Thus, (7) is the trivial result wherein for every coding scheme, the distortion increases if the rate decreases. Another property of rate-distortion curves is convexity. This is a required property because it implies that the parts of the spectrum with the least relevance to the entire signal are discarded first. A function is convex if the second derivative is monotonically decreasing. From (7), we obtain

$$\frac{\partial^2 D}{\partial R^2} = -\frac{\pi}{R_0^2}S'\left(\frac{R}{R_0}\pi\right). \tag{8}$$

We see that if the power spectral density function $S(\omega)$ is monotonically decreasing with increasing ω (i.e., $S'(\omega) < 0$), then the rate-distortion function is convex. Thus, in subsampling schemes, the convexity of the rate-distortion curve is directly coupled to the decreasing character of $S(\omega)$. If the power spectrum density is nondecreasing, the rate-distortion curve may become nonconvex.

If the argument presented in this section is extended to two dimensions, then after prefiltering and subsampling, $S(\omega_x, \omega_y)$

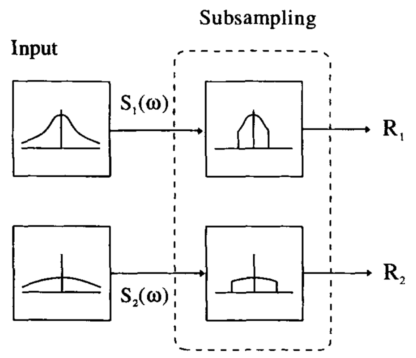


Fig. 3. Model for spatially adaptive subsampling.

does not extend over a specific bandwidth but covers a *region* U_L in the 2-D space. This region is the unity cell prescribed by the subsampling lattice. Equations (2) and (5) now become

$$R = \frac{\text{Area}(U_L)}{(2\pi)^2} R_0 \tag{9}$$

$$D = \sigma^2 - \frac{1}{(2\pi)^2} \int\limits_{(\omega_x,\omega_y)\in U_L}\int S(\omega_x,\omega_y) d\omega_x d\omega_y.$$

Without any specific knowledge about the 2-D subsampling lattice, it is not possible to obtain a close form relation for the rate-distortion function.

III. Spatially Adaptive Subsampling

A. A Model of a Spatially Adaptive Subsampling Scheme

In order to examine the advantage of spatially adaptive subsampling over fixed subsampling, a simple model is introduced. A block diagram of the model is given in Fig. 3. To account for the nonstationary nature of the input signal, we consider a signal that consists of two distinct regions (e.g., blocks in an image) with different statistics so that the power spectral density functions $S_1(\omega)$ and $S_2(\omega)$ differ from each other. We assume that the power spectral density functions exist and are monotonically decreasing; therefore, the corresponding rate-distortion functions are convex. Both regions have the same number of samples, and all samples are fed into a single coding scheme. The resulting bit rate R_t is given by the average of the two bit rates R_1 and R_2 used to code each region. The resulting distortion $D_t(R_t)$ after coding is given by

$$D_t(R_t) = D_1(R_1) + D_2(R_2) \tag{10}$$

where $D_1(R_1)$ and $D_2(R_2)$ are the rate-distortion functions of each region separately.

Let us first consider the case of fixed subsampling. Because both regions are encoded with the same sampling frequency, the following relation holds for the bit rates:

$$R_1 = R_2 = R_t. \tag{11}$$

The total distortion $D_t^F(R_t)$ can be computed by inserting this relation into (10):

$$D_t^F(R_t) = D_1(R_t) + D_2(R_t). \tag{12}$$

If spatially adaptive subsampling is used, the bits can be divided over the two regions in an uneven manner. In an optimal bit allocation where the mean square error is minimized, the resulting bit rate for each region is controlled by the following two relations:

$$R_t = R_1/2 + R_2/2 \tag{13}$$

and

$$\min_{R_1,R_2} D_1(R_1) + D_2(R_2) = D_t^A(R_t) \tag{14}$$

where $D_t^A(R_t)$ is the distortion in a spatially adaptive coding scheme. We now show in the following that if an optimal bit allocation is used, D_t^A is always less than or equal to D_t^F. Using the transformation $R_1/2 - R_2/2 = \Delta r$, where $|\Delta r|$ represents the difference between the allocated bit rate assigned to each region and the desired bit rate, (13) can be inserted into (14):

$$\min_{\Delta r} D_t^A(R_t) = \min_{\Delta r} D_1(R_t + \Delta r) + D_2(R_t - \Delta r). \tag{15}$$

The optimal bit allocation can now be solved by setting the first derivative with respect to Δr to zero:

$$\frac{\partial D_1(R_t+\Delta r)}{\partial \Delta r} + \frac{\partial D_2(R_t-\Delta r)}{\partial \Delta r} = 0. \tag{16}$$

Inserting (7) into this expression yields the following expression for the optimal bit allocation:

$$S_1((R_t+\Delta r)\pi) = S_2((R_t-\Delta r)\pi). \tag{17}$$

If $S_1(\pi R_t) > S_2(\pi R_t)$, then Δr is positive, meaning that more bits are assigned to the first region. If the power spectral density function of the first region is greater, then (7) implies that the slope of the rate-distortion curve of the first region is greater than that of the second region. This leads to

$$D_1(R_t) - D_1(R_t+\Delta r) > D_2(R_t-\Delta r) - D_2(R_t) \tag{18}$$

where we have used the convexity of the rate-distortion function. This relation indicates that the gain in assigning more bits to the first region is larger than the loss caused by assigning fewer bits to the second region. Rewriting (18) gives

$$\begin{aligned} D_1(R_t) + D_2(R_t) &> D_1(R_t+\Delta r) + D_2(R_t-\Delta r) \\ \Leftrightarrow D_t^F(R_t) &> D_t^A(R_t). \end{aligned} \tag{19}$$

The above argument also holds if the characteristics of the first and second regions are exchanged. If the power densities differ, then the performance of spatially adaptive subsampling is always better than the performance of nonadaptive subsampling. If the solution to (17) yields $\Delta r = 0$, then $S_1(\pi R_t)$ is equal to $S_2(\pi R_t)$. In this case, the performance of fixed subsampling is obviously identical to the performance of spatially adaptive subsampling.

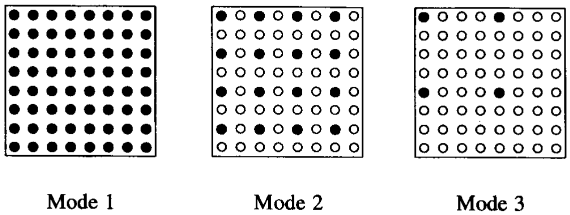


Fig. 4. Examples of different modes. The solid dots are the pixels that are transmitted.

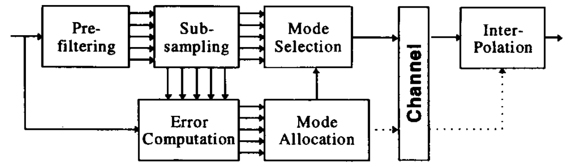


Fig. 5. General spatially adaptive coding scheme.

B. Spatial Adaptive Subsampling in Practice

Now that we have shown that spatially adaptive subsampling works better that fixed lattice subsampling from a theoretical point of view, we consider the practical implementation. The ideal case would be to segment the image into regions that require the same spatial sampling frequency and sample each region according to this frequency. Such a solution would require a detailed analysis of the image, and a large amount of side information would be needed to transmit the shape of the regions. Therefore, we subdivide the image into square blocks, and within each block, one specific sampling lattice is used. The size of the blocks is an important system parameter. If large blocks are chosen, the amount of side information is low, but the ability to adapt to the local spatial frequency contents would be lost. Small blocks cause a large overhead but warrant a better adaptation.

Another consideration in a practical system is the sampling lattice to be used for each block. Ideally, each block should be sampled with a sampling lattice optimally suited for that particular block. Again, this implies a large amount of side information. Therefore, only a limited set of possible sampling lattices is used. This set is designed in such a way that it gives a good coverage of the range of all the necessary spatial frequencies. In the sequel, each specific sampling lattice is called a *mode*. In Fig. 4, some examples of modes are given with different data reduction factors. For instance, in mode 3, only four pixels are kept out of 64 pixels, giving a data reduction factor of 16. Mode 1 can be used for highly detailed regions, whereas mode 3 can be used for areas with a slowly varying luminance. The number of possible modes is affected by the block size because for decreasing block size, the number of possible sampling lattices within the block decreases as well.

In a constant bit rate application, the output of the spatially adaptive subsampling scheme should be a fixed number of samples. Thus, the available modes should be distributed over the different blocks in such a way that the weighted sum of all the modes is equal to the desired bit rate. Of course, the image quality should be the best possible. Therefore, a criterion function that reflects the quality of the block for a particular mode needs to be used. In the next section, an allocation algorithm is discussed for this optimal mode distribution problem.

The resulting overall coding scheme is shown in Fig. 5. The input image is first prefiltered and subsampled for each mode. The subsampled images are fed into an interpolation module that also evaluates the quality criterion. The quality of each mode is used in the mode allocation that assigns a particular mode to each block. Finally, this information is transmitted to the receiver together with the pixels remaining after subsampling all blocks.

If, at the receiver, each block is interpolated using a technique that involves only the pixels within the block, the interpolation of each block is straightforward. However, if a more sophisticated interpolation technique is used, such as a filtering, information from neighboring blocks is required. This poses a problem if a neighboring block is sampled with a different mode because the pixels necessary for the interpolation may not be available. To avoid such problems, a hierarchical set of modes should be used. In a hierarchical set mode, $n+1$ is always a subset of mode n:

$$\{\boldsymbol{x} \mid \boldsymbol{x} \in \text{mode}_{n+1}\} \subset \{\boldsymbol{x} \mid \boldsymbol{x} \in \text{mode}_n\}, \quad \forall n \tag{20}$$

where the vector $\mathbf{x}$ represents a pixel location. For instance, the modes given in Fig. 4 form an hierarchical set. The mode with the smallest sampling density can be now be interpolated because the required boundary pixels are always present in the neighboring blocks. After interpolating all blocks with this mode, the same argument holds for the next mode in the hierarchy. It is obvious that this interpolation scheme will result in a worse interpolation result as compared with the interpolation made at the transmitter where all the required pixels are always present in the neighboring blocks. The severity of this loss of performance is examined in Section V. If a nonhierarchical set is used, a common intermediate sampling lattice must be introduced. From this lattice, the boundary pixels for the different modes can be deduced, and interpolation is still possible.

C. The Mode Allocation Problem

The mode allocation is of great significance as it influences the output quality considerably. A brute force search tries all the modes on all the blocks and selects the combination that gives the smallest distortion. In a practical situation, this is not feasible because in a system with N different modes and M blocks, the number of possible allocations is equal to N^M. As the number of blocks grows linearly, the number of allocations increases exponentially.

In a system where only two modes are used, the mode allocation problem can be solved analytically [2]. If the fraction of blocks sampled with mode 1 at a rate of R_1 is equal to α_1 and the fraction of blocks sampled with mode 2 at a rate of R_2 is α_2 ($R_1 > R_2$) then the following relations hold:

$$\alpha_1 + \alpha_2 = 1$$
$$\alpha_1 R_1 + \alpha_2 R_2 = R_t \tag{21}$$

where R_t is the desired total bit rate. From these equations, α_1 and α_2 can be solved. To minimize the distortion, all the blocks are first assigned to mode 2. Next, a fraction α_1 of the blocks with the highest distortion are assigned to mode 1. This will lead to an optimal mode assignment in the sense of minimizing the overall distortion.

If the number of modes is greater than two, (21) generalizes to a pair of equations with more than two unknowns, which cannot be solved uniquely. In [3], a heuristic algorithm is presented that does not guarantee optimality. Further, this algorithm is only suitable for three modes and is based on the 2-D histogram of the error differences between the different modes. A variation of this algorithm was described in [4]. Here, we describe a mode allocation algorithm that allows for an arbitrary number of modes and is under some conditions optimal.

The algorithm is based on the convex hull bit allocation scheme used for assigning quantizers to the subbands in a subband coding scheme [6]. Two modifications have to be made to this algorithm:

- The sources that have to be coded correspond to the blocks into which the image is subdivided.
- The quantizers correspond to the different mode structures.

Two assumptions made in the original algorithm require special attention. First, it is necessary that the rate-distortion curve of each source is convex. In Section II, we saw that this is not always the case and depends on the shape of the power spectral density function of each block. For a whole image, this may be a valid assumption, but for a small block in an image, this will no longer hold in general. A solution to this problem is to remove the modes that cause the rate-distortion curve to be nonconvex from the allocation process. The impact of this is that some allocations are no longer possible. If the optimal solution is contained in this set of excluded allocations, then the mode allocation in no longer optimal. The second requirement is that the total distortion is equal to the sum of the distortion for each source:

$$D_t = \sum_{i=1}^{M} D_i(R_i). \tag{22}$$

As we saw in the previous section, this condition cannot be satisfied because of the nonavailability of boundary pixels at the receiver. Hence, the mode allocation is performed on an *estimate* of the distortion. The effect of this problem is discussed in Section V.

The mode allocation algorithm starts by assigning each block $i(i \in \{1 \ldots M\})$ the mode with the lowest possible bit rate R_{0i}. This point gives the highest distortion, and it is guaranteed that this point lies on the rate-distortion curve. Starting from this point, for each block i, the rate difference ΔR_{ij} equal to

$$\Delta R_{ij} = |R_{0i} - R_{ij}| \tag{23}$$

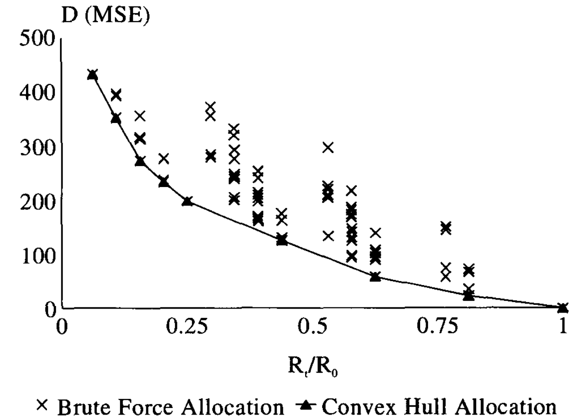


Fig. 6. Evaluated points in a mode allocation with four blocks and three different modes.

and the distortion difference ΔD_{ij}

$$\Delta D_{ij} = |D(R_{0i}) - D(R_{ij})| \tag{24}$$

of assigning mode $j(j \in \{1 \ldots N\})$ to that source is computed. The source with the smallest relative distortion gain ($\min_{i,j} \Delta D_{ij}/\Delta R_{ij}$) is assigned a new mode. In [6], it is proved that due to the convexity of the rate-distortion curve, this new mode allocation again lies on the rate-distortion curve. Starting from the new allocation, the described procedure is repeated. The algorithm terminates when the desired bit rate is reached. Hence, instead of evaluating all the possible mode allocations, only those lying on the rate-distortion function are examined. This causes a drastic decrease of the number of iterations and guarantees optimality. An example of the points evaluated is given in Fig. 6 for a system with $M = 4$ different blocks and $N = 3$ different modes different modes. In this case, a brute force allocation requires $3^4 = 81$ iterations. The fundamental benefit of the proposed search algorithm is to use an iterative method, which automatically follows the nine combinations being optimal from the rate-distortion point of view (solid triangles in Fig. 6).

IV. Motion-Compensated Spatially Adaptive Subsampling

Motion compensation has been used in many coding schemes to exploit the temporal correlation in image sequences. If a correct motion estimate is made, then no additional information has to be transmitted except for the motion vectors. A spatially adaptive subsampling scheme can benefit from this property if, for a particular region, the spatial correlation is low but the temporal correlation is high. In this case, spatially adaptive subsampling would require a lot of samples, whereas motion compensation requires none. The overall system is shown in Fig. 7. The shaded areas contain the components that were also present in the basic spatially adaptive subsampling scheme as shown in Fig. 5.

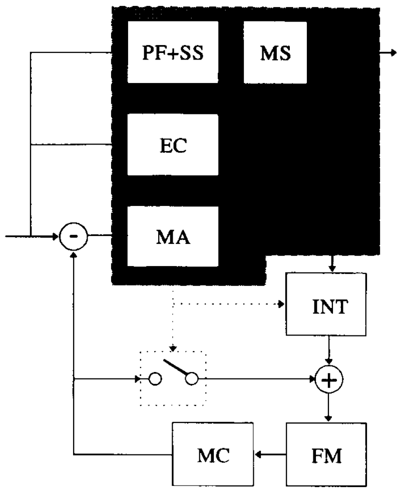


Fig. 7. Motion-compensated spatially adaptive coding scheme: (PF + SS is the prefilter and subsampling, MS is the mode selection, EC is the error computation, MA is the mode allocation, INT is the interpolation, FM is the frame memory, and MC is the motion compensation).

A motion-compensated prediction (MC) of the actual image is made using the previous image stored in the frame memory (FM). The prediction error is determined by subtracting the original image from the motion-compensated predicted image. In many coding schemes, the prediction error is coded and transmitted over the channel. Generally, this is not a good solution. The reason for this is that the computation of the prediction error acts as a high-pass filter and introduces extra high-frequency components. If an incorrect motion estimate is made and the spatial correlation is high (e.g., flat areas in the image), then the prediction error is still low. The area pointed to by the bad motion estimate can still be a reasonable prediction of the actual block. However, if a bad motion estimate is made and the spatial correlation is high (e.g., an area containing edges), extra high frequency components are introduced. This is because besides the edges in the original image, the edges of the previous image are also present in the prediction error. For a coding scheme, this property is not advantageous because of the decrease in spatial correlation. Thus, motion-compensated prediction can both reduce and increase the spatial correlation of the prediction error.

Therefore, a hybrid solution is chosen. Both the prediction error and the original image are used in the spatially adaptive coding scheme. The original image is subsampled (PF+SS) and the interpolation error is computed (EC) for the different modes. The prediction error is fed together with the interpolation errors into the mode allocation (MA). The mode allocation now starts by assigning to each block a mode with zero pixels. The interpolation for this mode is based on the predicted image. If the prediction error is high, then the modes used for spatially subsampling are assigned to the blocks in a similar manner as pure spatially adaptive subsampling. Hence, both the spatial and temporal correlation are exploited.

An advantage of this scheme is that it implicitly adjusts the threshold for the decision between intra and interframe coding. If the desired bit rate is high, then the algorithm is biased in the direction of intraframe coding, and if the bit rate is low, then interframe coding is preferred. Another advantage of this scheme is that now, there is a mode that requires no additional pixels; therefore, the maximal compression factor is no longer bounded by the block size, as is the case when only spatially adaptive subsampling is used.

V. Experiment Results

A. The Effect of the System Parameters

In this section, the effect of the different system parameters is investigated for spatially adaptive subsampling. This is illustrated by a practical example using the Lena test image. As a reference in all experiments, unless otherwise stated, a system with the mode structure as given in Fig. 4, a block size of 8 by 8 and a 3–tap prefilter is used. Mode information is accounted for in all experiments by reserving $\log_2(N)$ bits per block if N different modes are used. The transmission of the motion vectors is not accounted for. The algorithm is evaluated using the signal-to-noise ratio (SNR) defined by

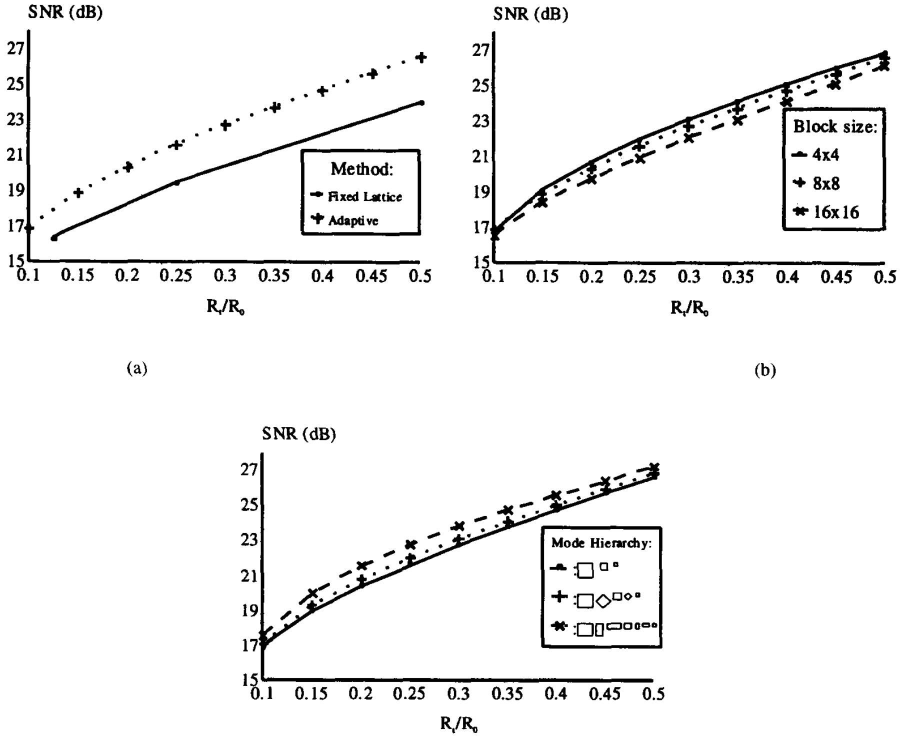
$$\mathrm{SNR} = 10^{10}\log\left(\frac{\sigma^2}{\mathrm{MSE}}\right) \tag{25}$$

where MSE is the mean square error.

First, spatially adaptive subsampling is compared with fixed lattice subsampling in Fig. 8(a). For the fixed lattice subsampling, horizontal, vertical, and quincunx subsampling was used and each time the scheme that gave the best result was selected. We see that spatially adaptive subsampling gives a significant improvement over fixed subsampling. The difference decreases with decreasing rate because eventually, for zero rate, the mean square error is equal to the image variance in both coding schemes.

In Fig. 8(b), the effect of different block sizes is shown. Small blocks give a better performance than large blocks because of a better adaptation to the local spatial frequency contents. For low bit rates, the advantage is smaller because there is a relative increase in the amount of side information compared to large blocks. In Fig. 8(c), the effect of different mode structures are shown. The first scheme is the mode structure as shown in Fig. 4. The second scheme starts from a block containing all samples and each time quincunx subsampling is used on the previous block. The third scheme is a modification of Fig. 4. In between the existing modes, two extra modes are introduced. One mode is a horizontal subsampling of the previous mode, and the second extra mode is a vertical subsampling of the previous mode. Note that this is not a hierarchical set. If the number of different modes increases, the performance increases as well. This is because a better adaptation is possible and a closer match can be found between the necessary Nyquist frequency and the available sampling frequencies.

Two different mode allocation schemes were compared, namely, the histogram method as described in [3] and the proposed algorithm. The performance of the histogram allocation



(c)

Fig. 8. Simulation result of Lena: (a) Comparison with fixed subsampling; (b) different block sizes; (c) different mode schemes.

did not differ much from the optimal allocation. However, the proposed allocation scheme can benefit from the fact that increasing the number of modes has a positive influence on the performance. In addition, the difference between the expected error given by the mode allocation and the actual error after interpolation was investigated. This difference was small, and therefore, it may be concluded that the impact of the errors introduced by different sampling structures at block boundaries can be neglected.

B. System Comparison

In this section, the following coding schemes are compared with each other:

- *Spatially adaptive subsampling (SA-SS):* This scheme uses no motion information.
- *Motion-compensated coding 2 (MC/NSS):* In this scheme, there are only two modes: If a good prediction is made, then noting is transmitted, and the original image is transmitted if a bad prediction is made. Hence, this scheme uses only temporal and no spatial adaptivity.
- *Motion-compensated spatially adaptive subsampling (MC/SA-SS):* This scheme uses motion-compensated prediction and spatially adaptive subsampling as described in Section IV.

The simulations were done using 20 frames of the Bicycles sequence, which consists of various moving objects.

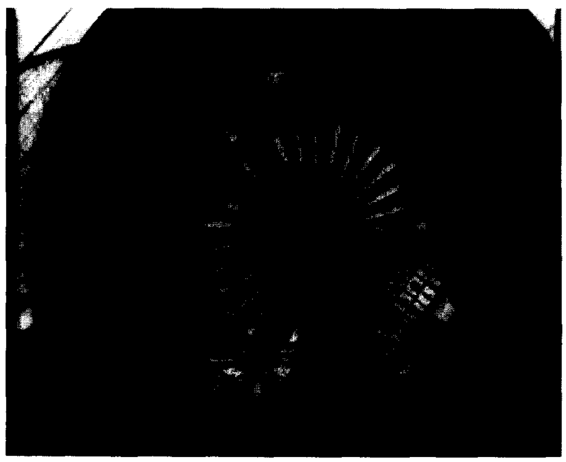


Fig. 9. Second image of original Bicycles sequence.

The second frame of the sequence is shown in Fig. 9. The size of the blocks was 4 by 4 pixels, and for the spatial adaptive subsampling, the hierarchy with seven modes from Fig. 8(c) was used. The motion vectors were estimated using hierarchical block matching with two levels. The results are shown in Fig. 10.

We see that spatially adaptive subsampling without motion compensation gives the worse performance. This is what

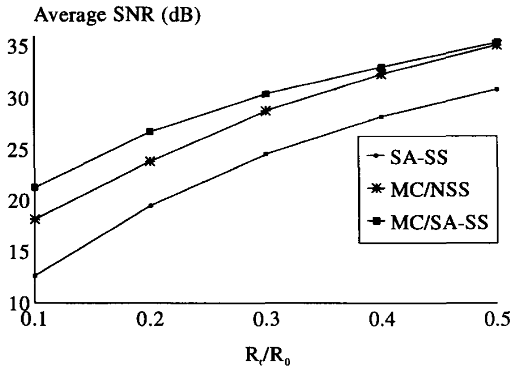


Fig. 10. Comparison of the different algorithms. The average SNR for the Bicyles sequence is shown.

Fig. 11. Mode assignment of second image of the Bicyles sequence at a reduction factor of 0.2. The grey level indicates the mode used: Dark indicates a high sampling density and light a low sampling density.

can be expected because no use is made of the temporal correlation in the image sequence. After motion compensation, the redundancy in the sequence is reduced considerably as is shown by the second scheme. Motion-compensated spatially adaptive subsampling gives the best coding result because both the spatial and temporal correlation can be exploited. The differences with the other coding results increases for low bit rates when the efficiency of the algorithm becomes more important. In Fig. 11, the mode assignment for the motion-compensated spatially adaptive subsampling scheme is shown. It can be observed that the regions with a constant luminance are assigned a mode with a low sampling density, whereas detailed regions are assigned a mode with a high sampling density.

VI. Conclusion

In this paper, spatially adaptive subsampling as an image coding method is described. Starting from a rate-distortion analysis of fixed lattice subsampling, it is proven that spatially adaptive subsampling gives a better performance in the rate-distortion sense compared with fixed lattice subsampling. The convex hull allocation algorithm based on rate-distortion theory is applied in modified form to assign the modes to the different blocks. We have seen that the major benefit of this allocation scheme compared with existing allocation schemes is that an arbitrary number of different modes can be used. This enables the extension of the coding system to three dimensions and gives a more flexible system than the existing spatially adaptive subsampling schemes.

In sub-Nyquist sampling schemes, which also use subsampling for bandwidth compression, only the temporal correlation of the image sequence is used. If the temporal correlation is strong, then a data reduction is achieved by spatially subsampling at the encoder and by temporal interpolation at the receiver. This is done regardless of the spatial content of the image sequence. However, motion-compensated spatially adaptive subsampling uses both the spatial and the temporal correlation in the image sequence. Therefore, the proposed scheme will work better than sub-Nyquist sampling schemes. Sub-Nyquist systems are usually incorporated into a digitally assisted television (DATV) system. Motion-compensated spatially adaptive subsampling can also be incorporated into a DATV system. The remaining pixels after subsampling can be transmitted using analog transmission. The necessary side information can be transmitted within the DATV channel.

References

[1] B. Girod and W. Geuen, "Vertical sampling rate decimation and line-offset decimation of colour difference signals," *Signal Processing*, vol. 16, no. 2, pp. 109–127, Feb. 1989.

[2] M. Tanimoto, N. Chiba, H. Yasui, and M. Murakami, "TAT (time-axis transform) bandwidth compression system of picture signals," *IEEE Trans. Commun.*, vol. 36, no. 3, pp. 347–354, Mar. 1988.

[3] R. Kishimoto and N. Sakurai, "A "high-efficiency TCM" bandwidth reduction for high-definition TV," in *Signal Processing HDTV: Proc. Second Workshop Signal Processing HDTV* (L. Chiariglione, Ed.). New York: Elsevier, 1988, pp. 129–136.

[4] M. Ashibe, K. Mitsuhshi, and S. Tsuruta, "A study on adaptive subsampling method for HDTV signal compression," in *Signal Processing HDTV: Proc. Second Int. Workshop HDTV* (L. Chiariglione, Ed.). New York: Elsevier, 1988, pp. 145–152.

[5] M. Tanimoto, A. Yamada, and K. Shibata, "A new TAT scheme for higher compression of HDTV," in *Signal Processing HDTV, II: Proc. Third Int. Workshop HDTV* (L. Chiariglione Ed.). New York: Elsevier, 1990, pp. 235–241.

[6] P. H. Westerink, J. Biemond, and D. E. Boekee, "An optimal bit allocation algorithm for sub-band coding," in *Proc. Int. Conf. Acoust., Speech Signal Processing*, Apr. 1988, pp. 757–760.

[7] F. W. P. Vreeswijk and M. R. Hagiri, "HDMAC coding for MAC compatible broadcasting of HDTV signals," in *Signal Processing of HDTV, II: Proc. Third Int. Workshop HDTV*. New York: Elsevier, Sept. 1989, pp. 187–194.

[8] R. A. F. Belfor, R. L. Lagendijk, and J. Biemond, "Subsampling of HDTV using motion information," in *Motion Analysis and Image Sequence Processing* (M. I. Sezan and R. L. Lagendijk, Eds.). Boston: Kluwer, 1993.

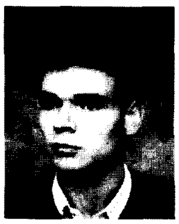
[9] E. Dubois, "The sampling and reconstruction of time-varying imagery with application in video systems," *Proc. IEEE*, vol. 73, no. 4, pp. 502–522, Apr. 1985.

[10] R. M. Mersereau and T. C. Speake, "The processing of periodically sampled multidimensional signals," *IEEE Trans. Acoustics, Speech Signal Processing*, vol. ASSP-31, no. 1, pp. 188–194, Feb. 1983.

[11] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.

Ricardo A. F. Belfor (M'92) was born in Zeist, The Netherlands, in 1966. He received the M.Sc. degree in electrical engineering from the Delft University of Technology, Delft, The Netherlands, in November 1989.

From 1990 to 1994, he worked at the Information Theory Group of the Delft University of Technology, where he did research towards the Ph.D. degree. His interests include image coding, motion estimation, and multidimensional digital signal processing.



Marc P. A. Hesp was born in Hilversum, The Netherlands, in 1968. He received the M.Sc. degree in electrical engineering from the Delft University of Technology, Delft, The Netherlands, in 1993.

His interests include digital signal processing.

Reginald L. Lagendijk (M'91) was born in Leiden, The Netherlands, in 1962. He received the M.Sc. and Ph.D. degrees in electrical engineering from the Technical University of Delft in 1985 and 1990, respectively.

In 1987, he became an Assistant Professor with the Laboratory for Information Theory of the Technical University of Delft. He was a Visiting Scientist in the Electronic Image Processing Laboratories of Eastman Kodak Research, Rochester, NY, in 1991. Since 1993, he has been an Associate Professor with the Laboratory for Information Theory of the Technical University of Delft. At present, his research interests include multidimensional signal processing and communication theory, with emphasis on filtering, compression, and analysis of image sequences. He is author of the book *Iterative Identification and Restoration of Images* (Kluwer, 1991) and co-author of the book *Motion Analysis and Image Sequence Processing* (Kluwer, 1993).

Dr. Lagendijk is an associate editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING.

Jan Biemond (F'92) was born in De Kaag, The Netherlands, on March 27, 1947. He received the M.S. and Ph.D degrees in electrical engineering from Delft University of Technology, Delft, The Netherlands, in 1973 and 1982, respectively.

He is currently Professor and Chairman of the Information Theory Group of the Department of Electrical Engineering at Delft University of Technology. His research interests include multidimensional signal processing, image enhancement and restoration, video compression (digital TV, stereoscopic TV, and HDTV), and motion estimation with applications in image coding and computer vision. He has authored and co-authored over 140 papers in these fields. In 1983 he was a Visiting Researcher at Rensselaer Polytechnic Institute, Troy, NY, and at Georgia Institute of Technology, Atlanta, GA.

Dr. Biemond is a member of the IEEE-SP Technical Committee on Image and Multidimensional Signal Processing and a member of the IEEE-CAS Technical Committee on Visual Signal Processing and Communications. He has served as the General Chairman of the Fifth ASSP/EURASIP Workshop on Multidimensional Signal Processing, which was held at Noordwijkerhout, The Netherlands, in September 1987. Further, he is an AdCom member of the European Association for Signal Processing (EURASIP) and a member of the Board of Governors of the IEEE Signal Processing Society since 1994. He is Co-Editor of the *International Journal on Multidimensional Systems and Signal Processing* and he serves on the Editorial Boards of *Image Communication*, and the *Journal of Visual Communication and Image Representation*. He is the scientific editor of a series of books on image communication. He was a Distinguished Lecturer of the IEEE Signal Processing Society for 1993-1994.