The Sampling and Reconstruction of Time-Varying Imagery with Application in Video Systems

ERIC DUBOIS, MEMBER, IEEE

Invited Paper

Sampling is a fundamental operation in all image communication systems. A time-varying image, which is a function of three independent variables, must be sampled in at least two dimensions for transmission over a one-dimensional analog communication channel, and in three dimensions for digital processing and transmission. At the receiver, the sampled image must be interpolated to reconstruct a continuous function of space and time. In imagery destined for human viewing, the visual system forms an integral part of the reconstruction process.

This paper presents an overview of the theory of sampling and reconstruction of multidimensional signals. The concept of sampling structures based on lattices is introduced. The important problem of conversion between different sampling structures is also treated. This theory is then applied to the sampling of time-varying imagery, including the role of the camera and display apertures, and the human visual system. Finally, a class of nonlinear interpolation algorithms which adapt to the motion in the scene is presented.

I. INTRODUCTION

Any image transmission system requires an initial sampling and reformatting operation which converts the original signal, in general a function of three independent variables (space and time), into a one-dimensional signal suitable for transmission over a communication channel. At the receiver an interpolation operation is carried out to convert the sampled signal back into a physically displayed image. In conventional analog television, the sampling is carried out in two dimensions only (vertical and temporal), by means of interlaced scanning. In digital processing and transmission systems, a full three-dimensional sampling is required.

A *conceptual* representation of an image sampling and reconstruction system is shown in Fig. 1. A time-varying scene is projected onto an image plane by an optical system, and a component such as luminance or a tristimu-

lus value is extracted to give a continuous function of space and time $u(x_1, x_2, t)$. This signal is filtered by a continuous three-dimensional low-pass filter and then sampled on a discrete set of points in space and time referred to as the sampling structure. The prefilter is required to suitably band-limit the input signal to avoid aliasing introduced by the sampling process. The sampled signal can then be digitally processed, stored, coded, etc. At the receiver, the signal must be interpolated to restore a continuous function of space and time for display and viewing. As will be seen, the operations in Fig. 1 cannot be neatly isolated in most practical systems. The general goal of a sampling system is to give the best possible rendition of the original image for a given spatiotemporal sampling density by appropriate choice of preprocessing, sampling structure, postprocessing, and interpolation. The response of the visual system should be considered when evaluating the performance of a system for the sampling and reconstruction of imagery destined for human viewing. A related problem is the conversion between different sampling structures in systems where more than one structure is used, or in interfacing different systems.

There has recently been intense activity in problems related to the sampling process, in order to provide higher quality pictures for the next generation of television systems. Examples are in high-definition television, extended-definition television, and camera and receiver processing for enhanced-quality television. At the opposite end of the spectrum, subsampling is a key technique in coding systems for very low data rates. These topics are all treated elsewhere in this special issue.

The goal of this paper is to present a general framework for the study of image sampling and interpolation, and to relate it to the specific application areas cited above. In Section II, the mathematical theory of sampling and interpolation of multidimensional signals is presented. The concept of sampling structure is introduced, and the Fourier representation and processing of signals defined on these structures is discussed. The multidimensional sampling the-
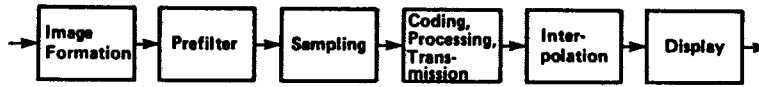
Fig. 1. Conceptual representation of an image sampling and reconstruction system.

orem is then presented, and finally processing for sampling structure conversion is treated. Section III then specializes this theory to the three-dimensional case of time-varying imagery. The role of the camera and display apertures and the human visual system in the sampling and reconstruction process are considered. Then structures suitable for image sampling are presented and evaluated. The sampling of color imagery is also discussed in this section. Techniques for interpolation are presented in Section IV. In particular, a special class of nonlinear interpolation schemes which adapt to the motion in the scene is described. These techniques recognize the special structure present in time-varying imagery due to the motion of objects, and also the differing resolution requirements of the human observer in moving and stationary areas.

II. Sampling of Multidimensional Signals

The basic operation is any image-sampling system is the specification of the image intensity or color on some regular array of points in space and time. The concept of a lattice, of importance in such areas as the geometry of numbers [1] and solid-state physics [2], is the basic tool in the study of image sampling. The theory of sampling multidimensional signals on a lattice was presented by Petersen and Middleton [3], and was later extended to periodically weighted sampling on a lattice [4]. A special case of periodically weighted sampling is sampling on a superposition of shifted lattices [5], [6]. This section presents the mathematical theory of sampled multidimensional signals, including the necessary results from the theory of lattices, Fourier transform representations, sampling of continuous signals, and conversion between different sampling structures. Image sampling has often been studied by representing it as the multiplication of the continuous signal with a regular array of Dirac delta functions. Although this is quite satisfactory for characterizing the sampling process alone, it makes the subsequent analysis of digital processing of the sampled signal difficult. This approach has thus been virtually abandoned in one-dimensional signal processing. In this paper, we adopt the approach that the sampled signal is truly discrete in space and time, with values only defined at the sample locations [7].

Although we are mainly concerned with the sampling of three-dimensional functions (i.e., time-varying two-dimensional images), the theory is easily developed for arbitrary dimension. Thus the theory in this section will be presented for the arbitrary multidimensional case, and will be equally applicable to still two-dimensional images, still three-dimensional images, and moving three-dimensional images, which are two-, three-, and four-dimensional signals, respectively. Section II-E presents a concise summary of the main results on lattices and sampling of continuous functions in three dimensions which is sufficient background for a first reading of Sections III and IV. Thus the application-oriented reader may wish to proceed directly to Section II-E on first reading.

A. Lattices

Definition [1]: Let $v_1, \cdots, v_D$ be linearly independent real vectors in D-dimensional Euclidean space R^D. A *lattice* Λ in R^D is the set of all linear combinations of $v_1, \cdots, v_D$ with integer coefficients

$$\Lambda = \{ n_1 v_1 + n_2 v_2 + \cdots + n_D v_D | n_i \in Z, i = 1, \cdots, D \}. \tag{1}$$

The set of vectors $v_1, \cdots, v_D$ is called a basis for Λ. A lattice is a discrete additive Abelian group. Fig. 2(a) shows an example of a lattice in two dimensions.

Let V be the matrix whose columns are the representation of the v_i with respect to the standard orthonormal basis for R^D. Then, the lattice is the set of all vectors Vn, with $n \in Z^D$. The basis for a given lattice is not unique. If E is any matrix of integers such that $\det E = \pm 1$, then the matrix $\hat{V} = EV$ provides another basis for Λ [1]. However, $|\det V|$ is unique and independent of the particular choice of basis. This quantity, denoted $d(\Lambda)$, is called the determinant of the lattice Λ and physically represents the reciprocal of the sampling density.

We define a *unit cell* of a lattice Λ as a set $\mathcal{P} \subset R^D$ (not necessarily connected) such that R^D is the disjoint union of copies of $\mathcal{P}$ centered on each lattice point: $(\mathcal{P} + x) \cap (\mathcal{P} + y) = \emptyset$ for $x, y \in \Lambda$, $x \neq y$, and $\bigcup_{x \in \Lambda}(\mathcal{P} + x) = R^D$. The hypervolume of a unit cell of a lattice Λ is $d(\Lambda)$. There are many possible choices for the unit cell of a lattice. One that is very convenient is the *fundamental parallelepiped* given by

$$\mathcal{P} = \left\{ \sum_{i=1}^{D} \alpha_i v_i | 0 \leqslant \alpha_i < 1 \right\} \tag{2}$$

where $v_1, \cdots, v_D$ is a basis for the lattice Λ. Another unit cell which is often useful is the *Voronoi cell* (also called Dirichlet region, Brillouin zone, Wigner–Seitz cell), the set of all points in R^D closer to 0 than to any other lattice point. Fig. 3 shows these two unit cells for the lattice of Fig. 2(a).

The concepts of sublattices and cosets of a lattice with respect to a sublattice are of importance in the theory of sampling on a superposition of shifted lattices, and conversion between sampling lattices.

Definition: Let Λ and Γ be lattices. Λ is a *sublattice* of Γ if every point of Λ is also a point of Γ. If Λ is a sublattice of Γ, then $d(\Lambda)$ is an integer multiple of $d(\Gamma)$. The quotient $d(\Lambda)/d(\Gamma)$ is called the index of Λ in Γ [1] and is denoted $(\Gamma:\Lambda)$. The set

$$c + \Lambda = \{ c + x | x \in \Lambda \} \tag{3}$$

for any $c \in \Gamma$ is called a *coset* or *class* of Λ in Γ. Two cosets are either identical or disjoint, and $c + \Lambda = d + \Lambda$ if and only if $c - d \in \Lambda$. There are $(\Gamma:\Lambda)$ distinct cosets of Λ in Γ, and the lattice Γ is the disjoint union of these $(\Gamma:\Lambda)$ cosets. A coset is a shifted version of the lattice Λ, and the set of cosets is the set of all shifted versions of Λ which are a subset of Γ.
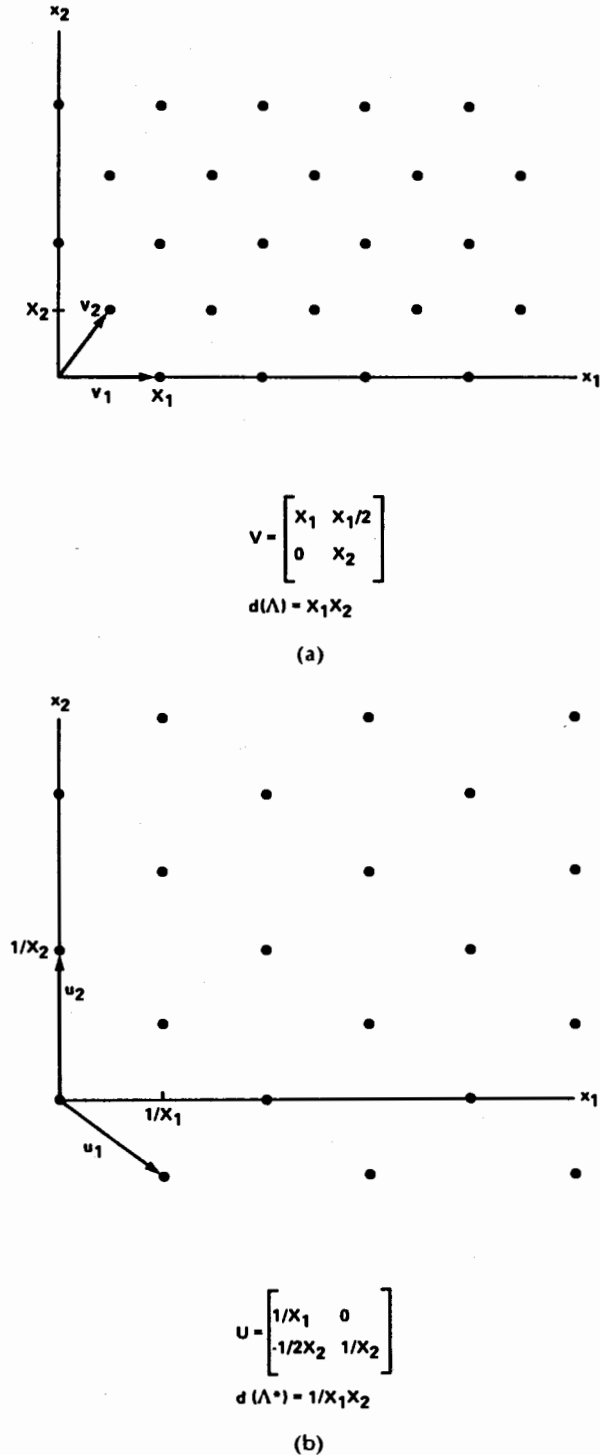
$$V = \begin{bmatrix} x_1 & x_1/2 \\ 0 & x_2 \end{bmatrix}$$

$$d(\Lambda) = x_1 x_2$$

(a)

$$U = \begin{bmatrix} 1/x_1 & 0 \\ -1/2x_2 & 1/x_2 \end{bmatrix}$$

$$d(\Lambda^*) = 1/x_1 x_2$$

(b)

Fig. 2. Example of a lattice in two dimensions. (a) Basic lattice Λ. (b) Reciprocal lattice Λ^*.

The intersection $\Lambda_1 \cap \Lambda_2$ of two lattices is also a lattice, although it is possible for the dimension of this lattice to be less than D. A necessary and sufficient condition for $\Lambda_1 \cap \Lambda_2$ to be of dimension D is that $V_1^{-1}V_2$ be a matrix of rational numbers, where V_1 and V_2 are the matrices for the lattices Λ_1 and Λ_2. The sum of two lattices $\Lambda_1 + \Lambda_2$ is defined as $\{x + y | x \in \Lambda_1, y \in \Lambda_2\}$. If $\Lambda_1 \cap \Lambda_2$ is a lattice of dimension D, then so is $\Lambda_1 + \Lambda_2$, and $(\Lambda_1 + \Lambda_2:\Lambda_1) = (\Lambda_2:\Lambda_1 \cap \Lambda_2)$. The intersection $\Lambda_1 \cap \Lambda_2$ is the largest lattice which is a sublattice of both Λ_1 and Λ_2, while the sum
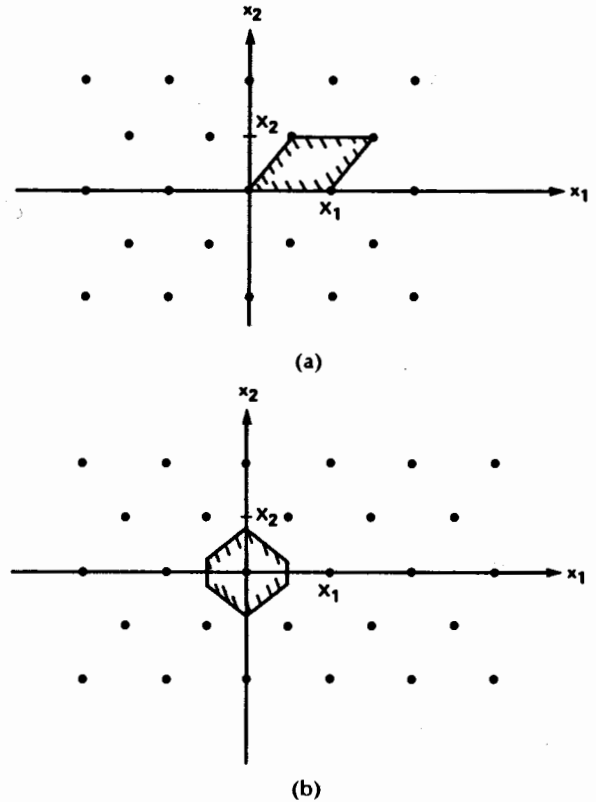

(a)

(b)

Fig. 3. Unit cells for the lattice of Fig. 2(a). (a) Fundamental parallelepiped. (b) Voronoi cell.

$\Lambda_1 + \Lambda_2$ is the smallest lattice which contains both Λ_1 and Λ_2 as sublattices.

A final concept which is very useful in the frequency-domain representation of signals sampled on a lattice is that of the reciprocal lattice.

Definition: Given a lattice Λ, the set of all vectors y such that $y^T x$ is an integer for all $x \in \Lambda$ is called the *reciprocal lattice* Λ^* of the lattice Λ.

The lattice Λ^* is also called the polar lattice in the geometry of numbers. A basis for Λ^* is the set of vectors $u_1, \cdots, u_D$ determined by $u_i^T v_j = \delta_{ij}, i, j = 1, \cdots, D$, or equivalently by $U^T V = I$ where I is a D by D identity matrix. The reciprocal lattice for the lattice of Fig. 2(a) is shown in Fig. 2(b). If Λ is a sublattice of Γ, then it is easily seen that Γ^* is a sublattice of Λ^*. Also, if Λ_1 and Λ_2 are lattices, it can be shown that $\Lambda_1 + \Lambda_2 = (\Lambda_1^* \cap \Lambda_2^*)^*$. Thus an algorithm for determining the intersection of lattices can also be used to determine the sum of lattices.

B. Multidimensional Sampled Signals

A sampling structure Ψ is a discrete set of points in R^D over which the image function is specified. The most general form of sampling structure considered in this paper is the union of selected cosets of a sublattice Λ in a lattice Γ. Thus we have

$$\Psi = \bigcup_{i=1}^{P} (c_i + \Lambda) \tag{4}$$

where $c_1, \cdots, c_P$ is a set of vectors in Γ such that $c_i - c_j \notin \Lambda$ for $i \neq j$. It is assumed that Γ is the smallest lattice containing Ψ and that there is no lattice Υ with $d(\Upsilon) < d(\Lambda)$ such

that Ψ is a union of cosets of Υ in Γ. By taking $\Lambda = \Gamma$ and $P = 1$, (4) reduces to a standard lattice. Fig. 4 shows an example of a sampling structure in two dimensions which cannot be represented as a lattice, but can be represented as the union of two shifted lattices. All sampling structures
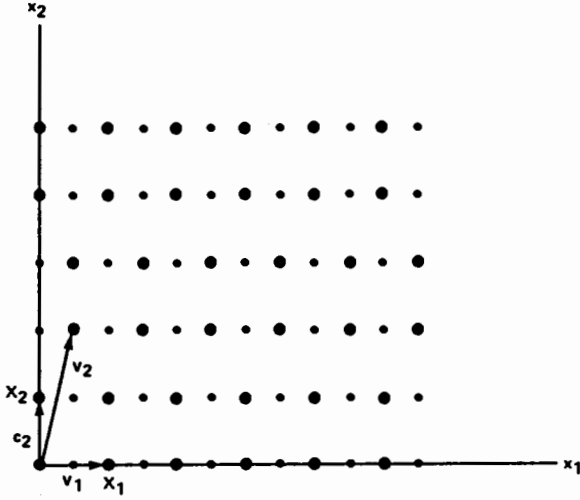

Fig. 4. Sampling structure in two dimensions which is the union of two cosets of the lattice Λ in the lattice Γ: $\Psi = \Lambda \cup (c_2 + \Lambda)$. v_1 and v_2 form a basis for Λ. Ψ is denoted by large dots and Γ by small dots.

known to the author which have been considered for image sampling can be represented in this way. Examples of sampling structures which are not lattices but can be represented using (4) can be found in [5], [6], [8].

The concept of a unit cell can usefully be extended to such a sampling structure. The unit cell is defined as a set $\mathscr{P} \subset R^D$ such that $\mathscr{P} \cap (\mathscr{P} + x) = \emptyset$ for $x \in \Psi$, $x \neq 0$, and $\bigcup_{x \in \Psi} (\mathscr{P} + x) = R^D$. For $P = 1$ this reduces to the standard definition.

We now consider the Fourier representation of signals sampled on such a structure. From Fourier analysis, the Fourier transform can be defined for any L^1 function defined on a discrete Abelian group [9]. Since a lattice is an Abelian group, we can define the Fourier transform over a lattice Λ

$$U(f) = \sum_{x \in \Lambda} u(x) \exp\left(-j2\pi f^T x\right)$$
$$= \sum_{n \in Z^D} u(Vn) \exp\left(-j2\pi f^T Vn\right), \quad f \in R^D. \quad (5)$$

The Fourier transform is a periodic function over R^D with periodicity lattice equal to the reciprocal lattice Λ^*, that is

$$U(f) = U(f + r), \quad \forall r \in \Lambda^*. \quad (6)$$

This follows from the fact that $r^T x$ is an integer for $x \in \Lambda$ and $r \in \Lambda^*$, by definition of the reciprocal lattice. Fig. 5 illustrates the periodicity of the Fourier transform for a signal defined on the lattice of Fig. 2. Because of this periodicity, the Fourier transform need only be specified over one unit cell $\mathscr{P}$ of the reciprocal lattice.

The inverse Fourier transform of the sampled signal is given by

$$u(x) = d(\Lambda) \int_{\mathscr{P}} U(f) \exp\left(j2\pi f^T x\right) df, \quad x \in \Lambda. \quad (7)$$
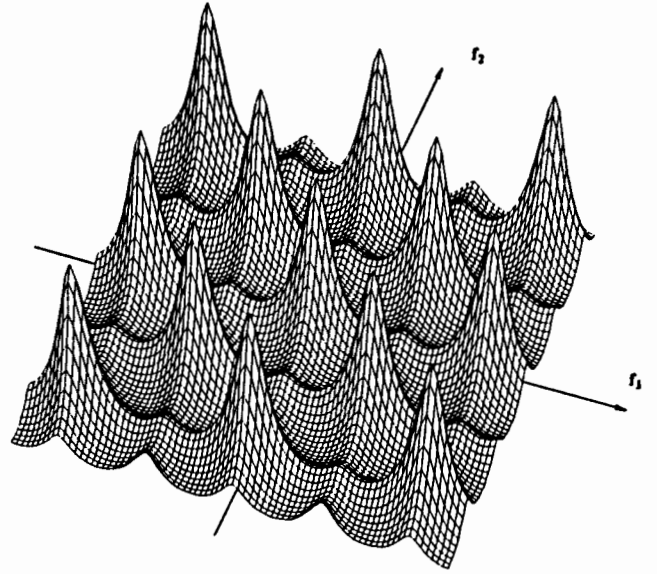

Fig. 5. Periodicity of Fourier transform for a signal defined on the lattice of Fig. 2(a). The periodicity is determined by the reciprocal lattice shown in Fig. 2(b).

For a sampling structure which is not a lattice, the Fourier transform in the usual sense is not defined. However, for the representation of the lattice as the union of certain cosets in a lattice Γ, we can assume that the signal is defined over the lattice Γ, with appropriate sample values set to zero. Then, the usual Fourier transform over Γ becomes

$$U(f) = \sum_{x \in \Psi} u(x) \exp\left(-j2\pi f^T x\right), \quad f \in R^D \quad (8)$$

and the Fourier transform will have periodicity determined by the reciprocal lattice Γ^*. For a sampling structure as given in (4), the Fourier transform can be written

$$U(f) = \sum_{i=1}^{P} \sum_{x \in \Lambda} u(c_i + x) \exp\left(-j2\pi f^T(c_i + x)\right)$$
$$= \sum_{i=1}^{P} \exp\left(-j2\pi f^T c_i\right) \sum_{x \in \Lambda} u(c_i + x) \exp\left(-j2\pi f^T x\right)$$
$$= \sum_{i=1}^{P} \exp\left(-j2\pi f^T c_i\right) U_i(f) \quad (9)$$

where $U_i(f)$ is the Fourier transform of the function $u_i(x) = u(c_i + x)$ defined on Λ.

In most cases, a sampled time-varying image cannot be considered as a function in $L^1(\Psi)$, but rather may more appropriately be considered as a sample from a discrete random field. For signals defined on a lattice Λ we assume in the usual way that the process is a homogeneous random field with zero mean and autocovariance function

$$R_u(x) = E[u(y)u(y + x)]. \quad (10)$$

The power density spectrum of the random field is given by the Fourier transform of the autocovariance function

$$\Phi_u(f) = \sum_{x \in \Lambda} R_u(x) \exp\left(-j2\pi f^T x\right). \quad (11)$$

A random field defined on the more general sampling structure of (4) cannot be homogeneous since the set of

points $\{x|y + x \in \Psi\}$ varies with y and thus $E[u(y)u(y + x)]$ cannot be independent of y. However, it can be periodic in y, with periodicity given by Λ, yielding a cyclostationary field. Then in the usual fashion [10], an autocovariance averaged over a period can be defined

$$R_u(x) = \frac{1}{(\Gamma:\Lambda)} \sum_{i=1}^{P} E[u(c_i)u(c_i + x)], \qquad x \in \Gamma. \tag{12}$$

The autocovariance function will be nonzero on the set

$$\mathcal{D} = \{x - y | x, y \in \Psi\}$$
$$= \bigcup_{i=1}^{P} \bigcup_{j=1}^{P} (c_i - c_j + \Lambda). \tag{13}$$

Clearly, not all of the P^2 cosets in this expression are distinct. The power density spectrum is then given by

$$\Phi_u(f) = \sum_{x \in \mathcal{D}} R_u(x) \exp(-j2\pi f^T x). \tag{14}$$

The theory of processing signals defined on a lattice has been presented in [11]. Of main interest to us is the case of linear filtering with finite-impulse response (FIR) filters. FIR filters are generally preferred over infinite-impulse response (IIR) filters for image processing because they allow exact linear phase response, and because stability is assured. For a linear shift-invariant system whose input and output are signals defined on a lattice Λ, the input and output are related by the convolution

$$z(x) = \sum_{y \in \Lambda} u(y)h(x - y), \qquad x \in \Lambda \tag{15}$$

where $h(x)$ is the unit sample response of the system. An FIR filter is characterized by the fact that $h(x)$ is nonzero for only a finite number of points $x \in \Lambda$. The frequency response of the filter is given by the Fourier transform of the unit sample response

$$H(f) = \sum_{x \in \Lambda} h(x) \exp(-j2\pi f^T x). \tag{16}$$

As before, the frequency response is periodic, with periodicity given by the reciprocal lattice Λ^*. If the input to an FIR filter is an L^1 function with Fourier transform $U(f)$, then the output is also an L^1 function with Fourier transform

$$Z(f) = H(f)U(f). \tag{17}$$

On the other hand, if the input to the filter is a homogeneous random field with power density spectrum $\Phi_u(f)$, then the output is also a homogeneous random field with power density spectrum

$$\Phi_z(f) = |H(f)|^2 \Phi_u(f). \tag{18}$$

A number of techniques exist for designing multidimensional FIR filters with frequency response approximating some desired characteristic. An overview of many of these is contained in [7]. The more general problem of linear processing when the input and output are defined on different lattices is considered in detail in Section II-D.

For signals not defined on a lattice, the ideas of shift invariance and frequency response of a linear filter are not straightforward, and this case is not considered here.

C. Sampling and Reconstruction

In general, multidimensional signals defined on a discrete sampling structure are obtained by sampling a continuous function over R^D. If this continuous function is denoted $u_c(x)$, $x \in R^D$, the operation of sampling on a structure Ψ is given by

$$u(x) = u_c(x), \qquad x \in \Psi. \tag{19}$$

Note that while u_c is defined over all of R^D, u is only defined on Ψ. The theory of sampling and reconstruction of both deterministic and random signals on a lattice was presented in [3]. In this section, we review this material, with the extension to structures of the form of (4).

Sampling: Suppose that $u_c \in L^1(R^D)$ has Fourier transform

$$U_c(f) = \int_{R^D} u_c(x) \exp(-j2\pi f^T x) \, dx, \qquad f \in R^D \tag{20}$$

with the inverse Fourier transform relation

$$u_c(x) = \int_{R^D} U_c(f) \exp(j2\pi f^T x) \, df, \qquad x \in R^D. \tag{21}$$

Then, this integral evaluated for $x \in \Lambda$ can be written as a sum of integrals over displaced versions of a unit cell $\mathcal{P}$ of Λ^*

$$u(x) = \int_{R^D} U_c(f) \exp(j2\pi f^T x) \, df$$
$$= \sum_{r \in \Lambda^*} \int_{\mathcal{P}} U_c(f + r)$$
$$\cdot \exp(j2\pi(f + r)^T x) \, df, \qquad x \in \Lambda. \tag{22}$$

By the property of the reciprocal lattice, $\exp(j2\pi r^T x) = 1$ so that exchanging the order of summation and integration gives

$$u(x) = \int_{\mathcal{P}} \left[\sum_{r \in \Lambda^*} U_c(f + r) \right] \exp(j2\pi f^T x) \, df. \tag{23}$$

Taking the Fourier transform of this gives

$$U(f) = \frac{1}{d(\Lambda)} \sum_{r \in \Lambda^*} U_c(f + r). \tag{24}$$

Thus the Fourier transform of the sampled signal is the sum of an infinite number of copies of the Fourier transform of the continuous signal, shifted according to the reciprocal lattice. This function has periodicity lattice Λ^* as required. Fig. 6 shows the effect of sampling with the lattice of Fig. 2 on the spectrum of a continuous band-limited signal. Examples where the shifted versions of the analog spectrum overlap and do not overlap are given.

If a homogeneous random field with autocovariance function $R_{uc}(x)$ and power density spectrum $\Phi_{uc}(f)$ is sampled on a lattice Λ, the autocovariance of the sampled signal is $R_u(x) = R_{uc}(x), x \in \Lambda$ so that the above development gives

$$\Phi_u(f) = \frac{1}{d(\Lambda)} \sum_{r \in \Lambda^*} \Phi_{uc}(f + r). \tag{25}$$

The situation for the sampling on a union of shifted lattices can be analyzed by combining the above results
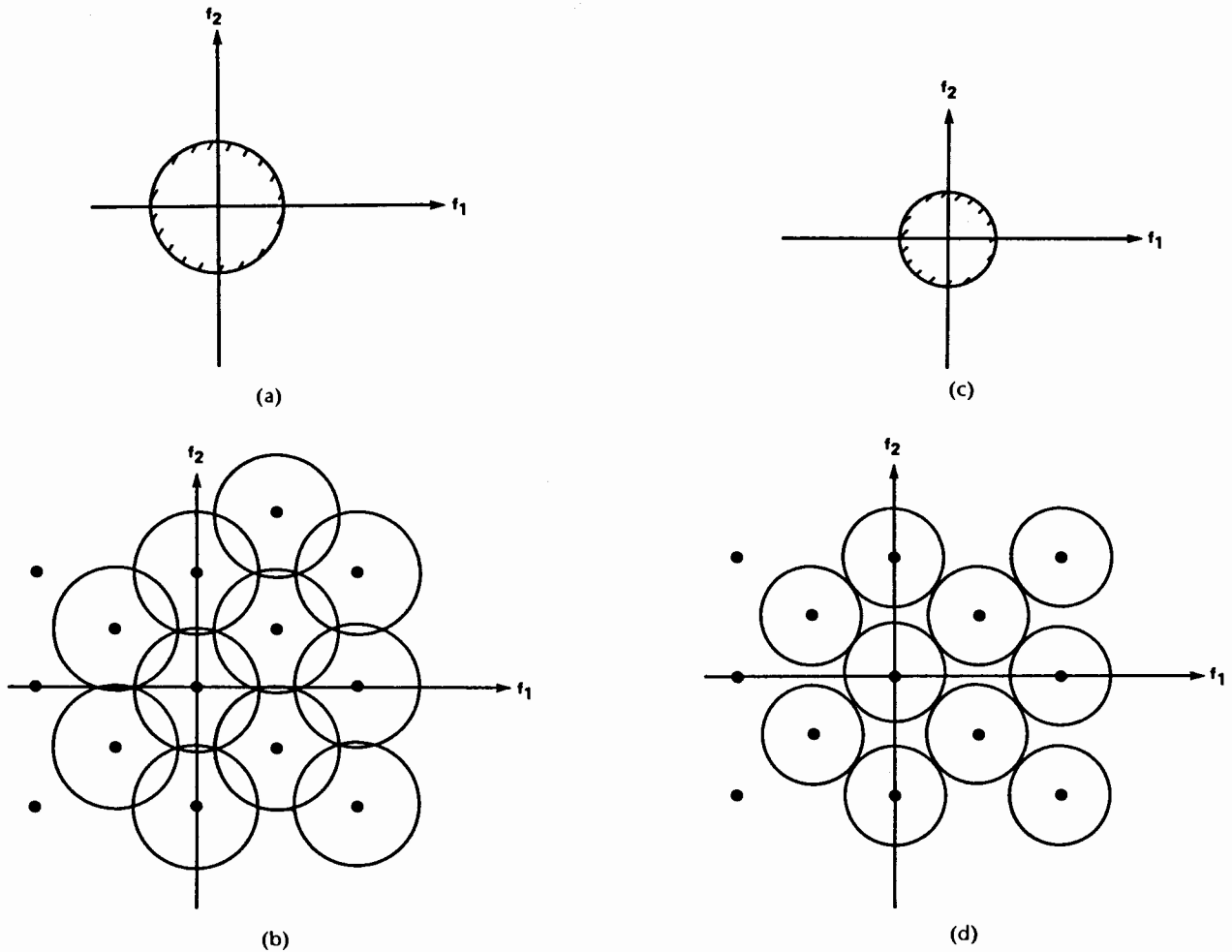
9

Fig. 6. Sampling of continuous signals. (a) Region of support of spectrum of continuous signal. (b) Spectrum of signal in (a) sampled with lattice of Fig. 2(a), showing overlap in repeated versions of basic spectrum. (c) Spectrum of continuous signal with smaller region of support than (a). (d) Spectrum of signal in (c) sampled with lattice of Fig. 2(a), showing no overlap in repeated versions of basic spectrum.

with (9)

$$U(f) = \frac{1}{d(\Lambda)} \sum_{i=1}^{P} \exp\left(-j2\pi f^T c_i\right) \sum_{r \in \Lambda^*} U_c(f + r)$$

$$\cdot \exp\left(j2\pi(f + r)^T c_i\right)$$

$$= \frac{1}{d(\Lambda)} \sum_{r \in \Lambda^*} \left(\sum_{i=1}^{P} \exp\left(j2\pi r^T c_i\right)\right) U_c(f + r)$$

$$= \frac{1}{d(\Lambda)} \sum_{r \in \Lambda^*} g(r) U_c(f + r). \qquad (26)$$

The function

$$g(r) = \sum_{i=1}^{P} \exp\left(j2\pi r^T c_i\right)$$

is constant over cosets of Γ^* in Λ^*, and may be equal to zero for some of these cosets, so that the corresponding shifted versions of the basic spectrum are not present. It is such a cancellation which would make this pattern of interest, rather than the less dense lattice Λ. Thus we

define Ψ^* to be the union of cosets of Γ^* in Λ^* for which $g(r)$ is nonzero. Fig. 7 shows the reciprocal structure Ψ^* for the structure Ψ of Fig. 4.
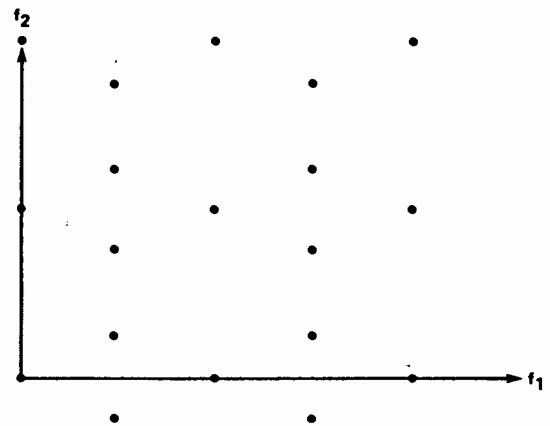

Fig. 7. Reciprocal structure to the sampling structure of Fig. 4.

10

Reconstruction: The final stage in an image communication system involves the reconstruction of the time-varying image as a continuous function of space and time for viewing. This reconstruction is an interpolation process which must fill in the missing values from the existing samples. Exact reconstruction of a continuous function from its samples on a structure Ψ is possible if the image spectrum is confined to a region $\mathcal{W} \in R^D$ such that the regions $\mathcal{W} + r$ for $r \in \Psi^*$ do not overlap with $\mathcal{W}$, i.e., there exists a unit cell $\mathcal{P}$ of Ψ^* such that $\mathcal{W} \subset \mathcal{P}$.[1] This is illustrated in Fig. 6(d). If the above conditions are satisfied, then

$$u_c(x) = \int_{\mathcal{W}} U_c(f) \exp\left(j2\pi f^T x\right) df$$

$$= \frac{d(\Lambda)}{P} \int_{\mathcal{W}} U(f) \exp\left(j2\pi f^T x\right) df. \quad (27)$$

Substituting the definition for $U(f)$ gives

$$u_c(x) = \sum_{s \in \Psi} u(s) t(x - s) \quad (28)$$

where

$$t(s) = \frac{d(\Lambda)}{P} \int_{\mathcal{W}} \exp\left(j2\pi f^T s\right) df. \quad (29)$$

If the smallest region of support of $U_c(f)$ is not a unit cell of the reciprocal structure, then the interpolating formula for $u_c(x)$ is not unique, since $\mathcal{W}$ can be chosen as any subset of a unit cell which contains the region of support. Equation (28) is a hybrid linear filtering, where the input is sampled and the output is continuous. The corresponding frequency response is

$$T(f) = \begin{cases} 1, & f \in \mathcal{W} \\ 0, & f \in \mathcal{W} + r, r \in \Psi^* - \{0\} \\ \text{arbitrary}, & \text{otherwise}. \end{cases} \quad (30)$$

This can in general be interpreted as an ideal low-pass filter. For homogeneous random fields, exact reconstruction (in the sense of vanishing mean-square error) is also obtained using (28) if the power density spectrum is limited to a unit cell of Λ^* [3].

If the signal is not band-limited to a unit cell, then a phenomenon known as aliasing occurs, whereby high frequencies in the original signal are mapped to lower frequencies. The most familiar examples are Moiré patterns, staircase effects on contours, and wagon wheels rotating backwards. A detailed discussion of the different types of aliasing effects which occur for a variety of sampling structures can be found in [6]. Petersen and Middleton [3] showed that if the original signal is not band-limited to a unit cell of the reciprocal lattice of the desired sampling lattice, the mean-square reconstruction error averaged over a cell of the sampling lattice can be minimized by prefiltering with a filter having unit gain over a suitably chosen unit cell, and zero gain elsewhere, again an ideal low-pass filter.

Partial Sampling: To date, all time-varying image recording or transmission systems use sampling in at least one

---

[1] If Ψ is not a lattice, exact reconstruction may be possible, even if the translated spectra overlap, by using a different interpolation function for each coset. See [12, p. 194] for a one-dimensional example.

---

dimension. Conventional analog television and cinema use only partial sampling, that is sampling in only one or two dimensions. Specifically, in analog television the signal is sampled in only two dimensions (essentially vertical and temporal), while in cinema the signal is sampled in the temporal direction only. We consider here the case of partially sampled signals, i.e., signals defined on a subset of R^D which is the direct sum of a lattice of dimension less than D and its orthogonal complement in R^D.

Let Λ be a C-dimensional lattice in R^D, generated by C linearly independent vectors $v_1, \cdots, v_C$ in R^D, where $C < D$. Let S_1 be the C-dimensional subspace of R^D spanned by $v_1, \cdots, v_C$, and let S_2 be the orthogonal complement of S_1 in R^D. A partially sampled signal defined on the set $\Psi = \Lambda + S_2$ is discrete in the dimensions determined by Λ and continuous in the dimensions determined by S_2. The set $\Lambda + S_2$ is an Abelian group, and the Fourier transform can be defined as

$$U(f_1, f_2) = \sum_{x \in \Lambda} \left(\int_{S_2} u(x, s) \exp\left(-j2\pi f_2^T s\right) ds \right)$$
$$\cdot \exp\left(-j2\pi f_1^T x\right) \quad (31)$$

where $f_1 \in S_1$ and $f_2 \in S_2$. If $u(x)$ is obtained by sampling a continuous signal $u_c(x)$ with Fourier transform $U_c(f)$, then

$$U(f) = \frac{1}{d(\Lambda)} \sum_{r \in \Lambda^*} U_c(f + r). \quad (32)$$

As an example, consider the following approximation to 2:1 interlaced scanning used in television. The lattice Λ is defined by the vectors $(0, 2X_2, 0)^T$ and $(0, X_2, X_3)^T$ and is simply a hexagonal lattice in the vertical–temporal plane given by $x_1 = 0$. The sampling structure is

$$R + \Lambda = \left\{ (x_1, 2n_2 X_2 + n_3 X_2, n_3 X_3)^T | x_1 \in R, n_1, n_2 \in Z \right\}.$$

D. Sampling Structure Conversion

It is often necessary to interface image communication systems which use different sampling structures. Some examples are conversion between the European and North American scanning standards, or conversion between different sampling structures used in video codecs. Another application of importance is in scan conversion between transmission and display scanning standards [13]. A complete treatment of sampling rate conversion for one-dimensional signals is given in [14], and a brief introduction to the problem of sampling structure conversion for multidimensional signals is found in [11]. This section presents the theory of linear filters with input and output defined on different sampling structures, as illustrated in Fig. 8. We assume that these sampling structures are lattices Λ_1 and Λ_2, and that the input and output signals are in $\ell^1(\Lambda_1)$ and
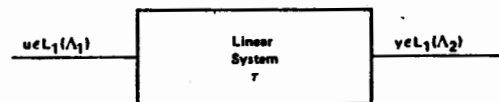

Fig. 8. Linear filter with input and output defined on different sampling lattices.

$l^1(\Lambda_2)$, respectively. The generalization to random fields is straightforward, but the generalization to nonlattice sampling structures is less so.

Let $\mathcal{T}$ be a linear system mapping functions defined on Λ_1 to functions on Λ_2, i.e., $\mathcal{T}:l^1(\Lambda_1) \rightarrow l^1(\Lambda_2)$. By linearity, we mean that $\mathcal{T}[a_1 u_1 + a_2 u_2] = a_1 \mathcal{T}[u_1] + a_2 \mathcal{T}[u_2]$ for any real numbers a_1 and a_2 and for any $u_1, u_2 \in l^1(\Lambda_1)$. Define the unit sample functions on a lattice by

$$\delta_s(x) = \begin{cases} 1, & x = s \\ 0, & x \neq s \end{cases} \qquad (33)$$

and the unit sample responses $h_s \in l^1(\Lambda_2)$

$$h_s = \mathcal{T}[\delta_s]. \qquad (34)$$

In the usual way, an arbitrary function can be written as a superposition of unit samples

$$u = \sum_{s \in \Lambda_1} u(s)\delta_s \qquad (35)$$

and by linearity, the output of the filter is given by

$$\mathcal{T}[u] = \sum_{s \in \Lambda_1} u(s)\mathcal{T}[\delta_s]$$

$$= \sum_{s \in \Lambda_1} u(s)h_s. \qquad (36)$$

A general property of interest in digital filters is shift invariance: if the input signal is shifted by p, the output signal is also shifted by p. In our case this is only possible if $p \in \Lambda_1 \cap \Lambda_2$. Thus we assume here that $\Lambda_1 \cap \Lambda_2$ is a D-dimensional lattice, i.e., that $V_1^{-1}V_2$ is a matrix of rational numbers. This is an extension of the condition that the ratio of sampling frequencies be a rational number in one-dimensional multirate systems [14]. Define the shift operator $\mathcal{S}_p$ on $l^1(\Lambda)$ by

$$\mathcal{S}_p u(s) = u(s - p). \qquad (37)$$

Then the required shift invariance property is

$$\mathcal{T}[\mathcal{S}_p u] = \mathcal{S}_p \mathcal{T}[u], \qquad p \in \Lambda_1 \cap \Lambda_2 \qquad (38)$$

or written out explicitly

$$\sum_{s \in \Lambda_1} u(s - p) h_s = \mathcal{S}_p \sum_{s \in \Lambda_1} u(s) h_s$$

$$= \sum_{s \in \Lambda_1} u(s) \mathcal{S}_p h_s. \qquad (39)$$

A simple change of variables on the left side gives

$$\sum_{s \in \Lambda_1} u(s) h_{s+p} = \sum_{s \in \Lambda_1} u(s) \mathcal{S}_p h_s. \qquad (40)$$

This equality will hold for arbitrary u if and only if

$$h_{s+p} = \mathcal{S}_p h_s, \qquad \forall p \in \Lambda_1 \cap \Lambda_2. \qquad (41)$$

From this equation we see that the unit sample responses over cosets of $\Lambda_1 \cap \Lambda_2$ in Λ_1 are translated versions of each other. Thus the maximum number of such unit sample responses required to completely specify the filter is equal to the index of $\Lambda_1 \cap \Lambda_2$ in Λ_1.

Let the index of $\Lambda_1 \cap \Lambda_2$ in Λ_2 be Q, and let $b_1, \cdots, b_Q$ be representatives for the cosets. If $y = \mathcal{T}[u]$, then the function y restricted to the ith coset can be written

$$y(b_i + p) = \sum_{s \in \Lambda_1} u(s) h_s(b_i + p)$$

$$= \sum_{s \in \Lambda_1} u(p - s) h_{p-s}(b_i + p)$$

$$= \sum_{s \in \Lambda_1} u(p - s) h_{-s}(b_i), \qquad p \in \Lambda_1 \cap \Lambda_2. \qquad (42)$$

This can be written in the more familiar convolution form

$$y(b_i + p) = \sum_{s \in \Lambda_1} u(p - s) f_i(s) \qquad (43)$$

where we define $f_i(s) = h_{-s}(b_i)$ for $i = 1, \cdots, Q$. This can be interpreted for each i as a linear filtering of u by the shift-invariant filter with impulse response f_i followed by a subsampling of the result to $\Lambda_1 \cap \Lambda_2$. The output y is obtained by multiplexing the output of the Q filters.

It is possible to represent this structure by a linear shift-invariant filter operating on a lattice which contains both Λ_1 and Λ_2. This is particularly useful for the application of frequency-domain filter design methods. Recall that $\Lambda_1 + \Lambda_2$ is the smallest lattice containing both Λ_1 and Λ_2. Define the upsampling operator $\mathcal{U}:l^1(\Lambda_1) \rightarrow l^1(\Lambda_1 + \Lambda_2)$ by

$$\mathcal{U}u(x) = \begin{cases} u(x), & x \in \Lambda_1 \\ 0, & x \notin \Lambda_1 \end{cases} \quad x \in \Lambda_1 + \Lambda_2 \qquad (44)$$

and the downsampling operator $\mathcal{D}:l^1(\Lambda_1 + \Lambda_2) \rightarrow l^1(\Lambda_2)$ by

$$\mathcal{D}v(x) = v(x), \qquad x \in \Lambda_2. \qquad (45)$$

Then the overall filter can be expressed

$$\mathcal{T} = \mathcal{D}\mathcal{T}^+\mathcal{U} \qquad (46)$$

where $\mathcal{T}^+:l^1(\Lambda_1 + \Lambda_2) \rightarrow l^1(\Lambda_1 + \Lambda_2)$ is a linear shift-invariant filter. This is illustrated in Fig. 9. The filtering operation is described by

$$v(x) = \sum_{s \in \Lambda_1 + \Lambda_2} w(s) h(x - s) \qquad (47)$$

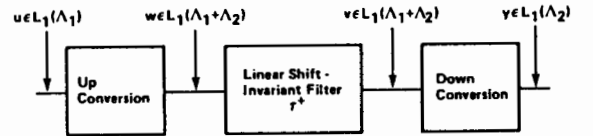where h is the unit sample response of $\mathcal{T}^+$. However, since

Fig. 9. Decomposition of sampling structure conversion system.

$w(x) = u(x)$ for $x \in \Lambda_1$, and is zero otherwise

$$v(x) = \sum_{s \in \Lambda_1} u(s) h(x - s), \qquad x \in \Lambda_1 + \Lambda_2. \qquad (48)$$

Finally, the downsampling operation gives

$$y(x) = \sum_{s \in \Lambda_1} u(s) h(x - s), \qquad x \in \Lambda_2. \qquad (49)$$

Comparing this with (36) gives

$$h(x - s) = h_s(x). \qquad (50)$$

It can easily be verified that this is a well-defined assignment. The functions f_i are obtained from h by

$$f_i(x) = h_{-x}(b_i) = h(b_i + x). \qquad (51)$$

A development very similar to that of Section II-B can be used to obtain the frequency-domain representation of the subsampling operation. Let $r_1, \cdots, r_N$ be coset representatives for the cosets of $(\Lambda_1 + \Lambda_2)^*$ in Λ_2^*, where $N = (\Lambda_1 : \Lambda_1 \cap \Lambda_2)$. Then

$$Y(f) = \frac{1}{N} \sum_{i=1}^{N} V(f + r_i). \qquad (52)$$

The overall filter is thus described by

$$Y(f) = \frac{1}{N} \sum_{i=1}^{N} H(f + r_i) U(f + r_i). \qquad (53)$$

When the change in sampling structure is such that overlap

in the replicated spectra is introduced (as in downsampling), part of the role of the filter $\mathscr{T}^+$ is to eliminate high-frequency signal components which contribute to this overlap. When the sampling density is increased, the filter serves to attenuate repeat spectra in a process of interpolation.

As an example of these ideas, consider a two-dimensional filter whose input and output are defined on the lattices Λ_1 and Λ_2 shown in Fig. 10(a) and (b). These lattices can be represented by the matrices

$$\begin{bmatrix} X_1 & 0 \\ X_2 & 2X_2 \end{bmatrix}$$

and

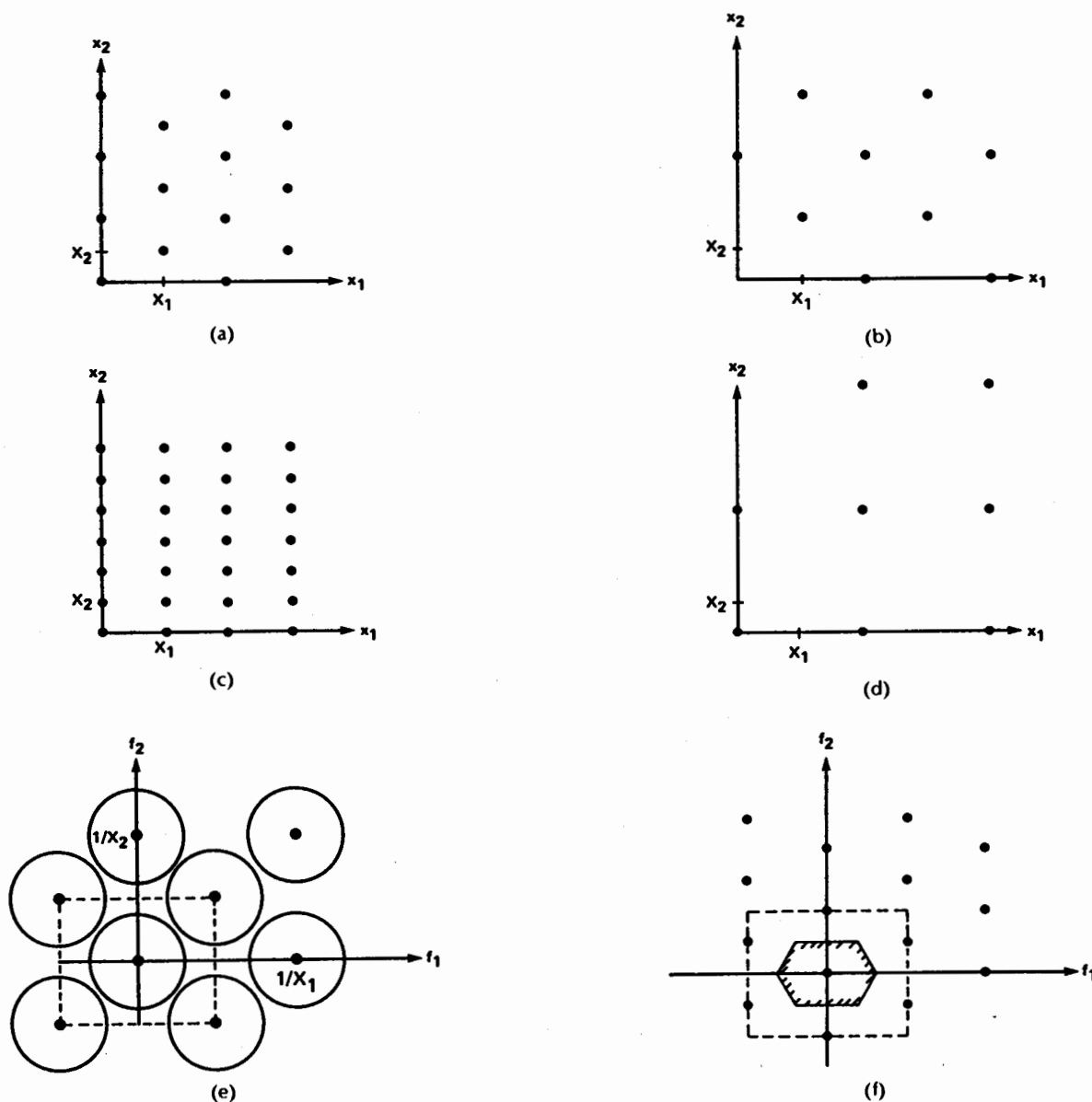$$\begin{bmatrix} X_1 & 0 \\ 2X_2 & 4X_2 \end{bmatrix}$$

Fig. 10. Example of sampling structure conversion. (a) Input lattice Λ_1. (b) Output lattice Λ_2. (c) $\Lambda_1 + \Lambda_2$. (d) $\Lambda_1 \cap \Lambda_2$. (e) Spectrum of input signal with periodicity Λ_1^*. (f) Passband of filter, equal to Voronoi cell of Λ_2^*. The dashed rectangle shows a basic period of the filter frequency response.

13

with $d(\Lambda_1) = 2X_1X_2$ and $d(\Lambda_2) = 4X_1X_2$. The sum $\Lambda_1 + \Lambda_2$ and intersection $\Lambda_1 \cap \Lambda_2$ of these lattices are shown in Fig. 10(c) and (d), respectively, and can be represented by the matrices

$$\begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}$$

and

$$\begin{bmatrix} 2X_1 & 0 \\ 0 & 4X_2 \end{bmatrix}.$$

Here, $Q = (\Lambda_1 + \Lambda_2 : \Lambda_1) = (\Lambda_2 : \Lambda_1 \cap \Lambda_2) = 2$. Suppose that the input signal has the spectrum shown in Fig. 10(e), with periodicity given by the reciprocal lattice Λ_1^*, whose matrix is

$$\begin{bmatrix} \dfrac{1}{X_1} & -\dfrac{1}{2X_1} \\ 0 & \dfrac{1}{2X_2} \end{bmatrix}.$$

The filter h defined on $\Lambda_1 + \Lambda_2$ has a frequency response which is periodic with respect to $(\Lambda_1 + \Lambda_2)^*$, and to avoid aliasing, should have a nonzero response only on a unit cell of Λ_2^*, as illustrated in Fig. 10(f). One period of $H(f)$ is indicated by the dashed rectangle in Fig. 10(e) and (f).

E. Summary of Three-Dimensional Sampling

This section highlights the main ideas of Section II, in particular as applied to three-dimensional sampling. A sampling lattice Λ for time-varying imagery is a discrete set of points in R^3 given by

$$\Lambda = \{ n_1 v_1 + n_2 v_2 + n_3 v_3 | n_1, n_2, n_3 \in Z \} \qquad (54)$$

where v_1, v_2, and v_3 are linearly independent vectors. The matrix $V = [v_1 | v_2 | v_3]$ is called the sampling matrix, and $d(\Lambda) = |\det V|$ is the reciprocal of the sampling density. The lattice determined by the vectors u_1, u_2, and u_3, where $U = [u_1 | u_2 | u_3] = (V^{-1})^T$ is called the reciprocal lattice Λ^*. Fig. 2 shows an example of a lattice and its reciprocal lattice.

The Fourier transform of a discrete signal defined on a lattice Λ is

$$U(f) = \sum_{x \in \Lambda} u(x) \exp(-j2\pi f^T x), \qquad f \in R^3.$$

The Fourier transform is periodic with respect to the reciprocal lattice

$$U(f) = U(f + r), \qquad r \in \Lambda^*.$$

One period of $U(f)$ is called a unit cell of the reciprocal lattice; there are many possible ways to define a unit cell (see Fig. 3 for two possible unit cells for a given lattice). If a continuous signal $u_c(x)$ with Fourier transform $U_c(f)$ is sampled on a lattice Λ

$$u(x) = u_c(x), \qquad x \in \Lambda$$

the Fourier transform of the sampled signal is

$$U(f) = \frac{1}{d(\Lambda)} \sum_{r \in \Lambda^*} U_c(f + r).$$

The spectrum of the sampled signal is thus the superposition of shifted versions of the original spectrum. Recon-struction of the continuous signal is in general possible if these do not overlap. The reconstruction filter is an ideal low-pass filter which passes the basic spectrum and eliminates the replicates. If overlap occurs, aliasing is said to take place; see Fig. 6 for an example. The total mean-square error can be minimized by prefiltering the input before sampling, to limit its spectrum to a unit cell of the recipro-cal lattice.

III. SAMPLING AND RECONSTRUCTION OF TIME-VARYING IMAGERY

The previous section has presented the mathematical theory of sampling of multidimensional signals. The trans-mission of a time-varying image over a one-dimensional channel requires the sampling of the image in at least two dimensions. If the original image is band-limited such that the sampling does not cause overlap of the repeated spec-tra, perfect reconstruction from the samples is possible. However, this is not the case in current television practice. There is aliasing in the sampling process, the reconstruction is far from perfect, and the human visual system is called upon to perform some of the interpolating postfiltering. This section discusses the issues related to sampling and reconstruction of time-varying imagery in television sys-tems, including scanning, sampling structures, and the role of the visual system.

A time-varying image is a function of three independent variables: horizontal and vertical spatial dimensions and time. This will be denoted $u_c(x) = u_c(x_1, x_2, x_3)$ where x_1 and x_2 are the horizontal and vertical coordinates, mea-sured in some convenient unit of length, and x_3 is time in seconds. The ultimate spatial unit of interest in imagery destined for human viewing is distance on the retina (or angle subtended at the eye) of the observer. However, for a given image, this depends on the size of the image display, and the distance of the viewer from the display. Since these quantities cannot in general be controlled, a unit of spatial distance related to the image size is usually used. In this paper the basic unit of spatial distance is the picture height (ph) with the corresponding spatial frequency unit of cycles per picture height (c/ph). If the distance from the viewer to the display in picture heights is known, these can be converted to degrees and cycles per degree subtended at the eye, respectively.

The sampling theory described in the previous section involves the use of ideal point sampling and ideal low-pass prefilters and interpolators. Real image sampling and dis-play involves the use of finite scanning apertures in the camera and display devices. The theory of scanning and display of still pictures with such apertures was first pre-sented by Mertz and Gray [15]. The extension of these ideas to three dimensions, using more convenient mathematical tools, has appeared more recently [16]–[19]. The following two sections develop these ideas within the framework we have established. Structures suitable for sampling time-vary-ing imagery are then discussed. Finally, the sampling of color video signals is treated.

A. Sampling

The process of ideal sampling as discussed in Section II-C requires the measurement of the image intensity at discrete

points or lines in space–time. This is not physically possible; a real sampling device measures the integral of the image intensity over a neighborhood of the desired point, weighted by an *aperture function a*, which may in general be space-variant. The sampled signal is thus

$$u(x) = \int_{R^3} u_c(x + s) a(x, s) \, ds, \qquad x \in \Psi. \qquad (55)$$

Under the assumption that the aperture is space-invariant, (55) becomes

$$u(x) = \int_{R^3} u_c(x + s) a(s) \, ds, \qquad x \in \Psi. \qquad (56)$$

It is clear from (56) that this is equivalent to the ideal sampling of the function

$$v(x) = \int_{R^3} u_c(x - s) h_a(s) \, ds, \qquad x \in R^3 \qquad (57)$$

where

$$h_a(s) = a(-s). \qquad (58)$$

In other words, this real sampling process is equivalent to the ideal sampling of the original image filtered by a linear continuous three-dimensional filter with frequency response

$$H_a(f) = \int_{R^3} h_a(s) \exp(-j2\pi f^T s) \, ds. \qquad (59)$$

The magnitude of $H_a(f)$ for $f_3 = 0$ is referred to as the modulation transfer function (MTF) of the camera.

The aperture function can normally be assumed to be separable in space and time. Although the spatial aperture is generally asymmetric and space-variant, it can be considered to a first approximation to be Gaussian. Fig. 11 shows a contour plot of the MTF for a circularly symmetric Gaussian aperture having a response of 0.5 at 400 TV lines (i.e., 200 c/ph), $H_a(f_1, f_2) = \exp(-(f_1^2 + f_2^2)/(240)^2)$. Miller [20] has given the MTF of several camera tubes, measured using the RCA P-300 test chart; Fig. 11 is typical of the
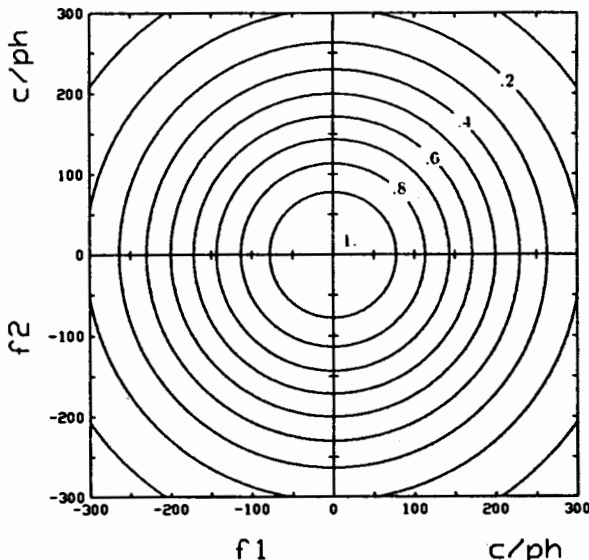
camera tubes described in [20]. The characteristics of the many camera types available are discussed elsewhere in this issue.

To get an idea of the temporal aperture of a camera, consider a tube which exactly integrates the light intensity over a time T and then is completely erased. The temporal aperture in this case is given by

$$h_a(t) = \begin{cases} 1/T, & 0 \le t \le T \\ 0, & \text{elsewhere.} \end{cases} \qquad (60)$$

Taking the Fourier transform of this gives the temporal response

$$H_a(f_3) = \exp(-j\pi f_3 T) \frac{\sin(\pi f_3 T)}{\pi f_3 T}. \qquad (61)$$

Fig. 12 shows a vertical–temporal slice (at $f_1 = 0$) of the magnitude frequency response of the resulting separable three-dimensional aperture for $T = 1/60$ s.
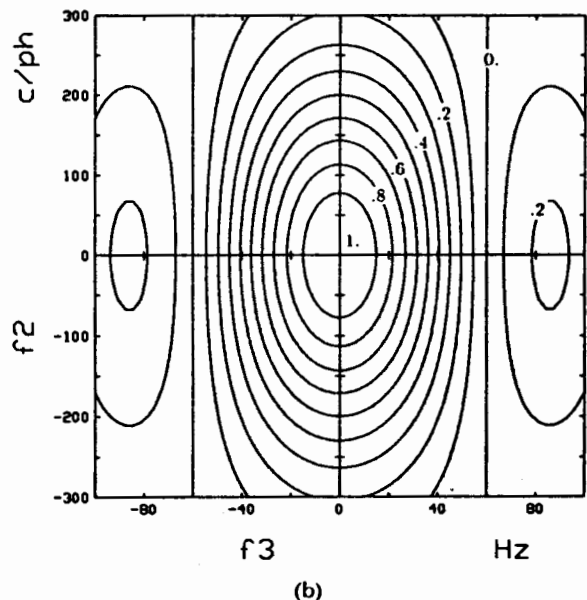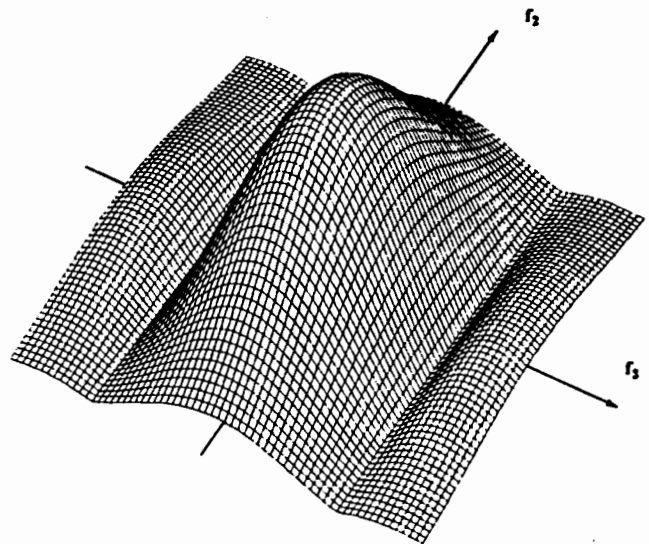

Fig. 11. Contour plot of frequency response of typical spatial Gaussian aperture. The response is 0.5 at 200 c/ph.

Fig. 12. Frequency response of typical vertical–temporal aperture. (a) Perspective view. (b) Contour plot.

The prefiltering provided by the camera aperture does not allow a good approximation to the ideal prefilter described in Section II-C, which passes all frequencies within a unit cell of the reciprocal lattice, and eliminates all frequencies outside the unit cell. There is a delicate balance between leaving high frequencies which will cause aliasing, and excessive attenuation of frequencies within the unit cell, which will compromise system resolution. For example, Schade [21] recommends that the camera MTF be no more than about 0.35 at half the vertical sampling rate to keep aliasing distortion at an acceptable level. However, this still results in significant attenuation of signal components below half the sampling rate, which could in principle be accurately maintained.

Although it is possible that better prefiltering could be obtained by electrooptical means, an alternative approach is to initially scan the image at a high sampling density, such that aliasing is negligible. The signal can then be converted to the desired sampling structure, performing the prefiltering digitally. It is also possible with this filter to compensate for the in-band signal attenuation due to the scanning aperture, a process known as aperture correction. With the use of digital prefilters, an arbitrarily sharp cutoff can be obtained. Although an infinitely sharp cutoff filter gives optimal results in terms of mean-square error, such a filter will cause ringing at sharp transitions, the well-known Gibbs' effect, which can be visually disturbing. A compromise must thus be struck between aliasing, loss of resolution, and ringing, to give optimal picture quality. A good illustration of these tradeoffs can be found in [22].

B. Reconstruction

The sampled image must be converted back to a continuous function of space and time for viewing on some display device such as a cathode-ray tube (CRT). The interpolation process given in (28) and (29) represents an ideal low-pass filtering. For human viewing, it is not necessary to perform perfect reconstruction, even in the ideal case. This is because signal distortions which are below the threshold of visibility can be allowed. In particular, since the visual system has a low-pass characteristic in spatial and temporal frequencies, the reconstruction processing need only reduce the magnitude of the high-order repeat spectra so that they are below threshold. However, since viewing conditions cannot generally be controlled, these thresholds should be based on the worst case viewing distance. Fig. 13 shows a perspective view of the threshold surface of the visual system as a function of spatial and temporal frequency, for particular viewing conditions [23]. Since the visual system is a nonlinear system, this figure can only serve as general indication as to the response to spatiotemporal patterns. A more complete model which accounts for the nonlinearity of the response and masking effects, such as that proposed by Lukas and Budrikis [24], would be required to fully predict the visibility of distortions. Note that the frequencies in Fig. 13 are at the retina, and may not correspond in a simple fashion to the spatiotemporal frequencies present in the imagery, if there is relative motion between the eye and the scene, as when objects are tracked. The relationship when the relative motion is uniform translation is derived in [6].
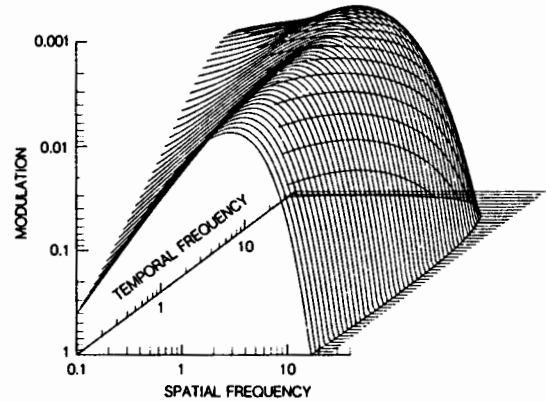

Fig. 13. Perspective view of portion of spatiotemporal threshold surface of human visual system [23]. Spatial frequency in c/deg, temporal frequency in Hz.

As in the camera, it is not easy to manipulate the display aperture to obtain a good approximation to the desired interpolating aperture. The reconstructed signal is given by

$$u_r(x) = \sum_{s \in \Psi} u(s) d(x - s) \qquad (62)$$

where $d(x)$ is the display aperture (impulse response). For a CRT display, the aperture is often assumed to be Gaussian, although more accurate approximations exist [21]. The temporal characteristic (phosphor decay) is usually exponential or logarithmic. The P22 phosphors used in color television decay to less than 10 percent of peak response in 10 μs to 1 ms [25]. Thus virtually no temporal filtering, and very little spatial filtering, is performed, in order to minimize loss of spatiotemporal resolution.

As a result of this imperfect interpolation, the viewer must be placed at a great enough distance for the visual filter to sufficiently attenuate the low-order repeat spectra. This may be at 6 to 8 times the picture height, as compared with a distance of 2 to 4 times which the viewer would naturally choose if no sampling artifacts were present [26]. This effect is illustrated by a rather extreme example of 18:1 subsampling shown in Fig. 14. At a normal reading distance of about 30 cm, the sampling structure obscures the text in Fig. 14(b), making reading difficult. By observing the figure from a greater distance (e.g., 1.5 m), the sampling structure is filtered by the visual system, and the text is more easily read. Fig. 14(c) shows the result of applying a two-dimensional interpolation filter, which attenuates the repeat spectra, making the text readable at 30 cm.

An approach to obtain an improved display aperture, similar to that described in Section III-A for the sampling aperture, can be taken. Digital interpolation to a higher scanning rate is performed, strongly attenuating the lowest order repeat spectra. The remaining repeat spectra are now in bands of lower visibility, and the gentle filtering of the display aperture is adequate to render them below threshold.

With current television practice, the combination of the camera aperture and the display aperture result in a maximum vertical resolution of only about 0.7 of the theoretical limit. This factor is sometimes referred to as the Kell factor, although there is no precise and universally accepted definition. An excellent discussion of the parameters of an
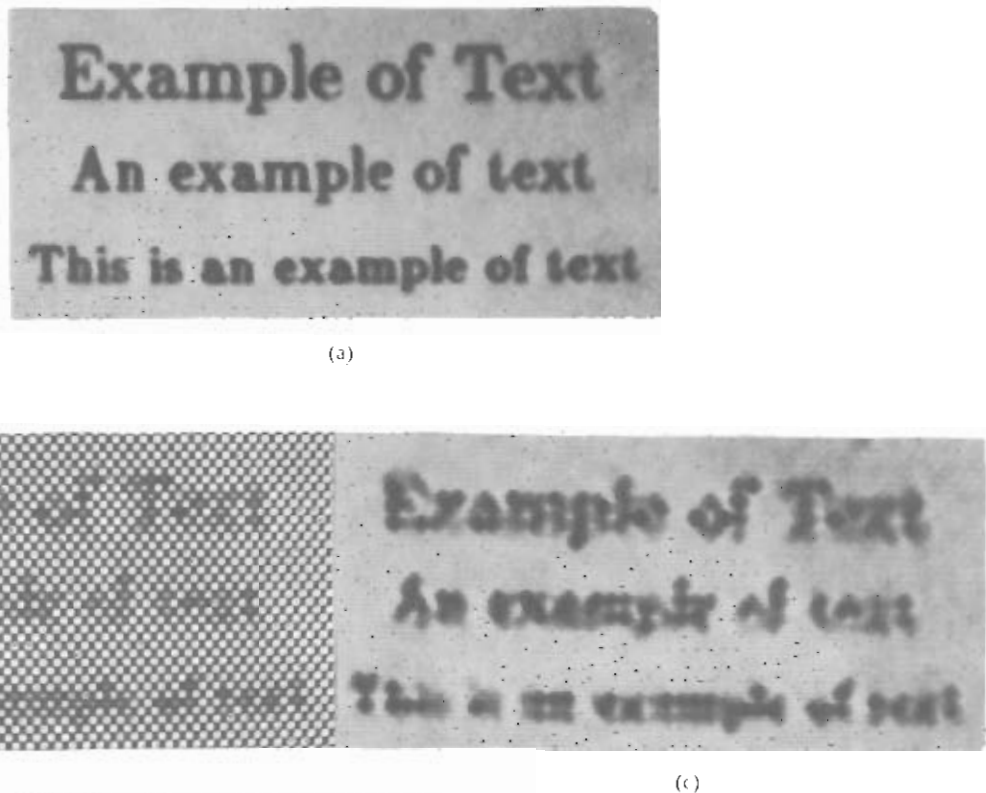
(a)

(b)

(c)

Fig. 14. Example of sampled text. (a) Original image. (b) Subsampled by a factor of 18 with hexagonal lattice and reconstructed with very narrow Gaussian aperture. (c) Reconstructed with two-dimensional interpolation filter.

imaging system which can influence the Kell factor has been given by Tonge [27]. By the approaches described above, significant improvement is possible, and Kell factors much closer to unity may be possible. This topic is discussed in more detail elsewhere in this issue.

C. Sampling Structures

Scanning: Scanning is the two-dimensional sampling process currently used in television, whereby the scene is sampled in the vertical and temporal directions, but is continuous in the horizontal direction. The resulting lines are then abutted to form a one-dimensional signal referred to as the video signal. Two vertical–temporal sampling structures are of interest: the orthogonal and the hexagonal structures shown in Fig. 15(a). These correspond to sequential and 2:1 interlaced scanning, respectively. The reciprocal lattices corresponding to these structures are shown in Fig. 15(b). As shown in Section II-C, the spectrum of the original image is replicated in the vertical–temporal frequency plane according to the reciprocal lattice.

The scanning lines are usually assumed to be perpendicular to the vertical–temporal plane. This is not precisely true; the lines have both a vertical and temporal tilt. For a 525-line sequential image with a 4:3 aspect ratio, this corresponds to rotating the image by about 0.16°. The temporal tilt means that the bottom of the image is scanned X_3 seconds after the top of the image, which could have some minor effect on large rapidly moving objects. However, as argued in [6] and [28], these effects are generally insignificant. We will ignore the tilt of the scanning lines in the remainder of this paper, and assume that they are

perpendicular to the vertical–temporal plane. The set of samples taken at a given time t constitute a *field*.

Both scanning patterns shown in Fig. 15 have the same number of lines per picture (counting both fields for the interlaced case) and the same sampling density. Examining the reciprocal lattices, we see that the closest replicated spectra to the origin for sequential scanning are at $(0, 1/X_2, 0)^T$, $(0, 0, 1/2X_3)^T$, and $(0, 1/X_2, 1/2X_3)^T$. For interlaced scanning, the closest replicated spectra to the origin are at $(0, 1/X_2, 0)^T$, $(0, 0, 1/X_3)^T$, and $(0, 1/2X_2, 1/2X_3)^T$. As mentioned previously, these replicated spectra are attenuated very little by the display aperture. The main degradation associated with the replicated spectrum at $(0, 1/X_2, 0)^T$ is visibility of the scanning lines; it is the same with both patterns. The main artifact associated with a component at $(0, 0, F)^T$ is large-area flicker at F hertz. With sequential scanning, large-area flicker is at $1/2X_3$ hertz, while for interlaced scanning, it is at $1/X_3$ hertz. If $X_3 = 1/60$ s (1/50 s), sequential scanning gives 30-Hz (25-Hz) large-area flicker, which is visually unacceptable. Interlaced scanning gives 60-Hz (50-Hz) large-area flicker which is significantly less visible, being almost imperceptible at 60 Hz, but perhaps slightly annoying at 50 Hz. The main distortions associated with the third components at $(0, 1/X_2, 1/X_3)^T$ or $(0, 1/2X_2, 1/X_3)^T$ are interline flicker and line crawl. These distortions are, of course, more visible with interlaced scanning than with sequential scanning, and are generally the most annoying defects in an interlaced display. However, they are still much less visible than the large-area flicker for which they have been traded. In general, 2:1 interlaced scanning display is preferable to sequential scanning for a given scanning density [26], [29].
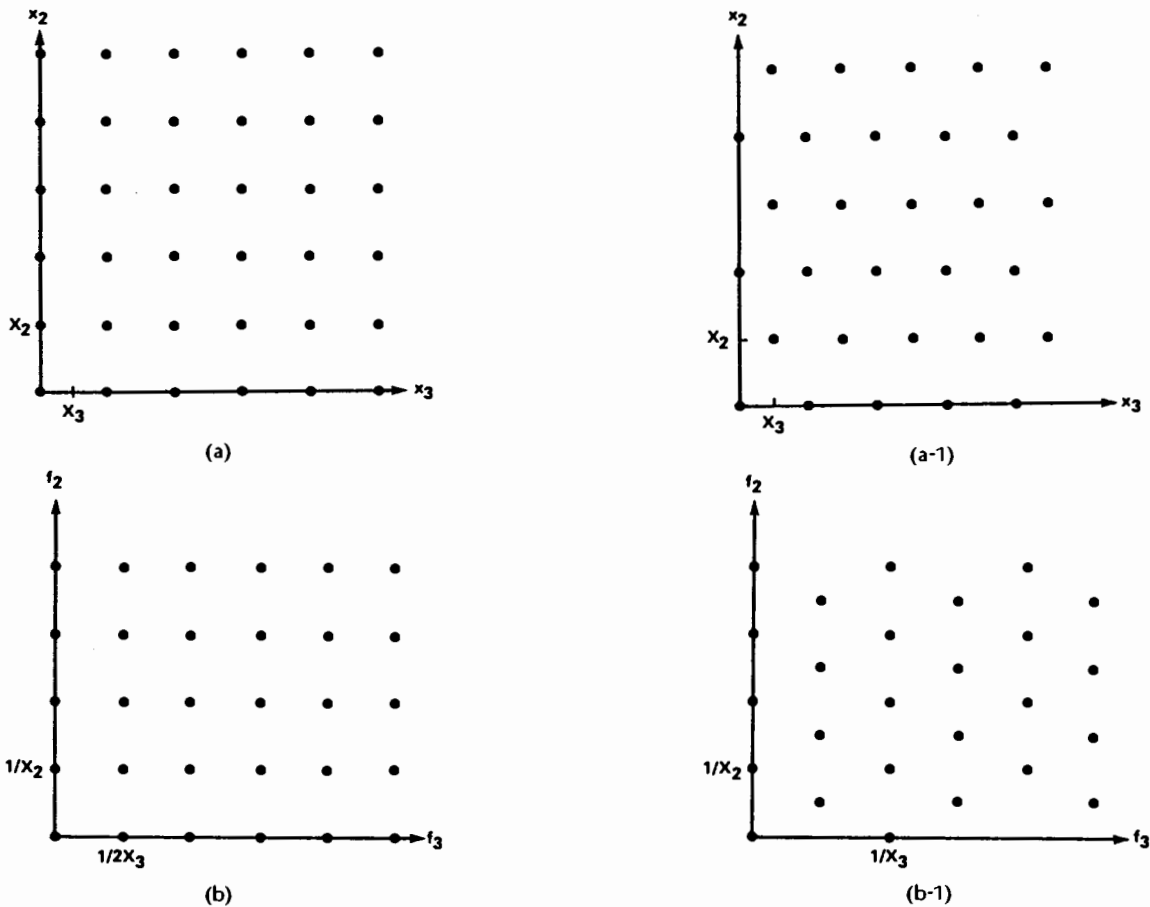
Fig. 15. (a) Two-dimensional sampling patterns for sequential and interlaced scanning of time-varying imagery. (b) Reciprocal lattices.

The spectrum of the one-dimensional video signal obtained by scanning can be related to the three-dimensional spectrum of the time-varying image. It can be shown that the one-dimensional spectrum is basically obtained by scanning the three-dimensional spectrum, giving the familiar comb-type spectrum. A detailed derivation can be found in [18].

Three-Dimensional Sampling Structures: For digital processing, coding, and transmission, the image signal must be sampled in three dimensions. This is usually done by horizontally sampling a signal which has already been scanned in the vertical–temporal dimension. However, with some solid-state sensors, the signal may be inherently sampled in three dimensions. Numerous sampling structures have been proposed for three-dimensional image sampling. A number of these will be presented in this section, and their relative merits discussed.

One of the earliest applications of a sampling lattice to television transmission was described by Gouriet [30] in 1952, in the context of dot-interlaced television. In the mid-1960s, Brainard *et al.* [31] carried out experiments with *replenishment patterns* for low-resolution TV; some were lattices while the rest were the union of two shifted lattices. In this work, no explicit interpolation was carried out, leaving this to the visual system. Further work on sampling structures came with the advent of digital television. Sabatier and Kretz [32] compared several structures for sampling of line-interlaced television signals and later presented an in-depth study of the performance of these sampling struc-

tures [6]. This work also emphasized interpolation by the human visual system. Tonge has described at length the design of three-dimensional digital filters for prefiltering and interpolation in two IBA research reports [28], [33].

Figs. 16–22 show a number of structures which have been proposed for image sampling. Also shown on these figures are the reciprocal structures and the lattice matrices. Sampling structures have often been illustrated by means of perspective drawings, as in [6] and [28]. However, for some of the more complex structures, these figures can be difficult to interpret, so we have chosen to show the spatial and spatial-frequency projections of the structures. All points in the structure occurring at the same time or temporal frequency carry the same number. The spacing in the temporal or temporal-frequency dimension between successive numbers is equal to the (3,3) element of the associated lattice matrix. For convenience, all the matrices are given in upper triangular form, so that the U matrices are not the inverse transposed of the V matrices, although the product UV^T is an integer matrix of determinant one. Fig. 23 shows a perspective view of the lattice and reciprocal lattice of Fig. 18, with the same numbering of the lattice points. Also shown is a Voronoi unit cell of the reciprocal lattice.

The closest lattice points to the origin in the reciprocal lattices indicate which frequencies in the input signal are most likely to cause aliasing problems. Kretz and Sabatier [6] have shown that the most critical structures for the sampling process are periodic structures and contours or
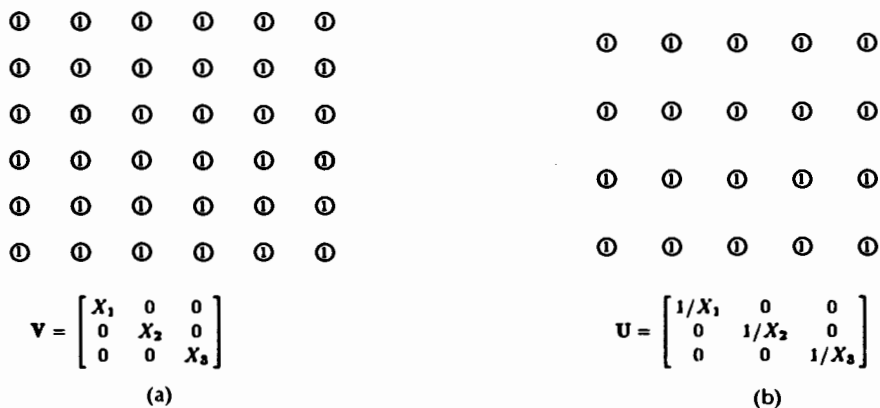
18

$$\mathbf{V} = \begin{bmatrix} X_1 & 0 & 0 \\ 0 & X_2 & 0 \\ 0 & 0 & X_3 \end{bmatrix}$$

(a)

$$\mathbf{U} = \begin{bmatrix} 1/X_1 & 0 & 0 \\ 0 & 1/X_2 & 0 \\ 0 & 0 & 1/X_3 \end{bmatrix}$$

(b)

Fig. 16. (a) Orthorhombic lattice (ORT). (b) Reciprocal lattice.

$$\mathbf{V} = \begin{bmatrix} X_1 & 0 & 0 \\ 0 & 2X_2 & X_2 \\ 0 & 0 & X_3/2 \end{bmatrix}$$

(a)

$$\mathbf{U} = \begin{bmatrix} 1/X_1 & 0 & 0 \\ 0 & 1/X_2 & 1/2X_2 \\ 0 & 0 & 1/X_3 \end{bmatrix}$$

(b)

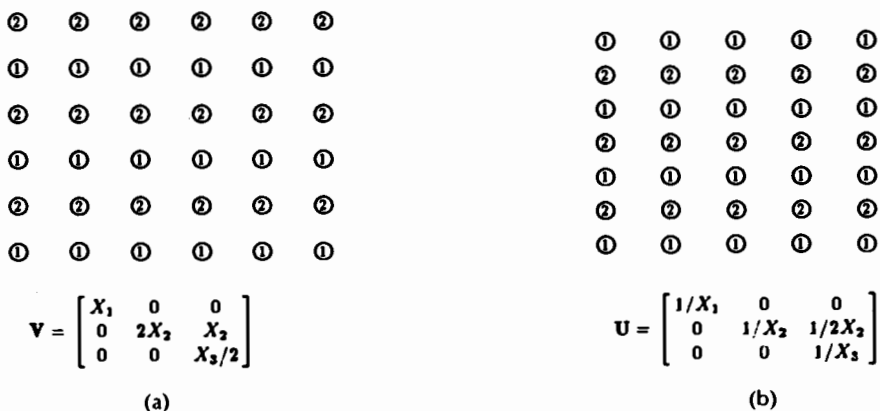**Fig. 17.** (a) Lattice obtained by vertically aligned sampling of 2:1 line-interlaced signal (ALI). (b) Reciprocal lattice.

$$\mathbf{V} = \begin{bmatrix} X_1 & 0 & X_1/2 \\ 0 & 2X_2 & X_2 \\ 0 & 0 & X_3/2 \end{bmatrix}$$

(a)

$$\mathbf{U} = \begin{bmatrix} 2/X_1 & 1/X_1 & 1/X_1 \\ 0 & 1/2X_2 & 0 \\ 0 & 0 & 1/X_3 \end{bmatrix}$$
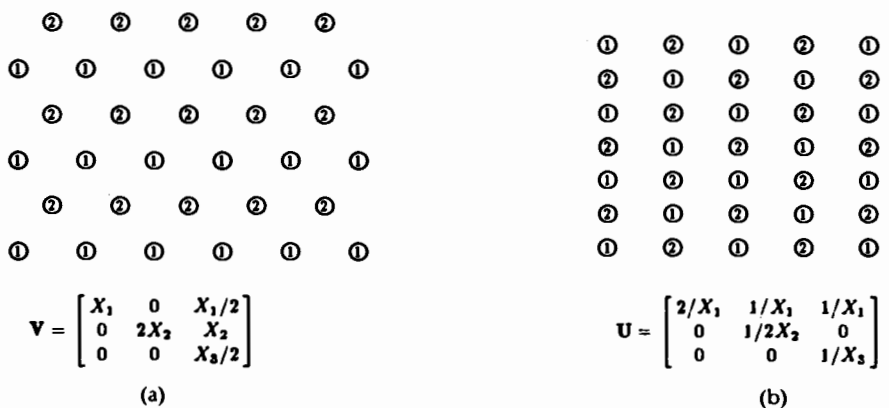
(b)

Fig. 18. (a) Body-centered orthorhombic lattice (BCO). Also known as field-quincunx (QT). (b) Reciprocal lattice.

edges. The distortion for a periodic structure is worst when any of its significant frequency components are near one of the lattice points of the reciprocal lattice. The distortion appears as a beat frequency at the difference between the structure frequency and the frequency of the closest reciprocal lattice point. When the aliasing is mainly spatial, this phenomenon is referred to as a Moiré pattern. For contours, the distortion appears as a phase perturbation along the length of the contour. This distortion is worst when the Fourier spectrum of the contour is oriented towards one of the points of the reciprocal lattice.

A good indication of the efficiency of a sampling lattice is

the ability to pack spheres densely without overlap on the points of the reciprocal lattice. This indicates the sampling density required to sample without aliasing a signal whose spectrum is confined to a spherical region in frequency space. Table 1 gives the minimum sampling density, and the corresponding values of X_1, X_2, and X_3, for each of the sampling structures in Figs. 16–22. The sampling density $C = 8W^3$ of the simple cubic lattice is used as a reference against which the others are compared. The optimum sampling density is 0.707C, which can be obtained with the lattices BCO, FCO, and HEX3. In all three cases, the reciprocal lattice is equivalent under rotation to the face-centered
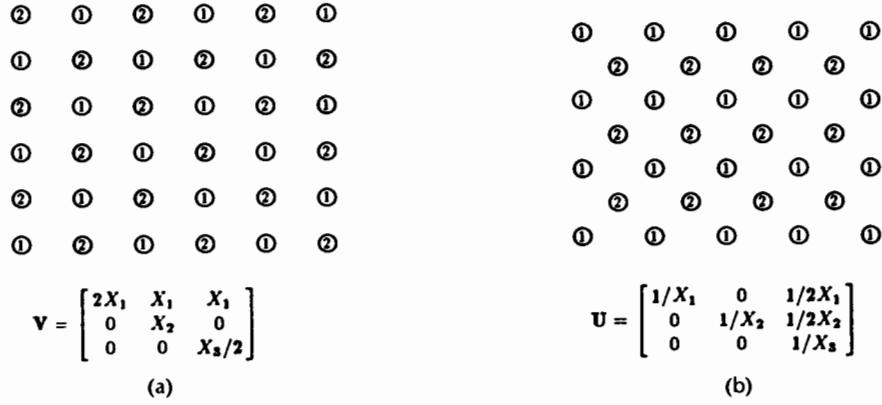
19

$$\mathbf{V} = \begin{bmatrix} 2X_1 & X_1 & X_1 \\ 0 & X_2 & 0 \\ 0 & 0 & X_3/2 \end{bmatrix}$$

(a)

$$\mathbf{U} = \begin{bmatrix} 1/X_1 & 0 & 1/2X_1 \\ 0 & 1/X_2 & 1/2X_2 \\ 0 & 0 & 1/X_3 \end{bmatrix}$$

(b)

Fig. 19. (a) Face-centered orthorhombic lattice (FCO). (b) Reciprocal lattice.

$$\mathbf{V} = \begin{bmatrix} 2X_1 & X_1 & 0 \\ 0 & 2X_2 & X_2 \\ 0 & 0 & X_3/4 \end{bmatrix}$$

(a)

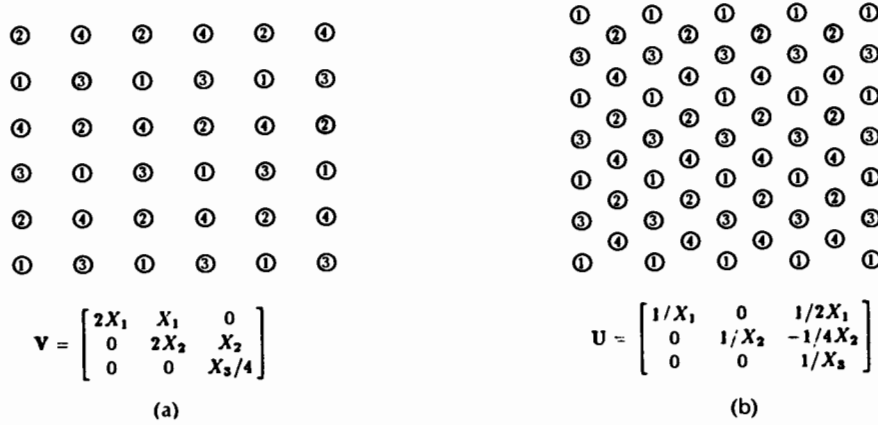$$\mathbf{U} = \begin{bmatrix} 1/X_1 & 0 & 1/2X_1 \\ 0 & 1/X_2 & -1/4X_2 \\ 0 & 0 & 1/X_3 \end{bmatrix}$$

(b)

Fig. 20. (a) Hexagonal lattice with four field periodicity (HEX4). Also known as nonstationary line-quincunx (QLNS). (b) Reciprocal lattice.

$$\mathbf{V} = \begin{bmatrix} 3X_1 & 3X_1/2 & X_1 \\ 0 & X_2 & 0 \\ 0 & 0 & X_3/3 \end{bmatrix}$$

(a)

$$\mathbf{U} = \begin{bmatrix} 2/X_1 & 1/X_1 & 2/3X_1 \\ 0 & 1/2X_2 & 0 \\ 0 & 0 & 1/X_3 \end{bmatrix}$$
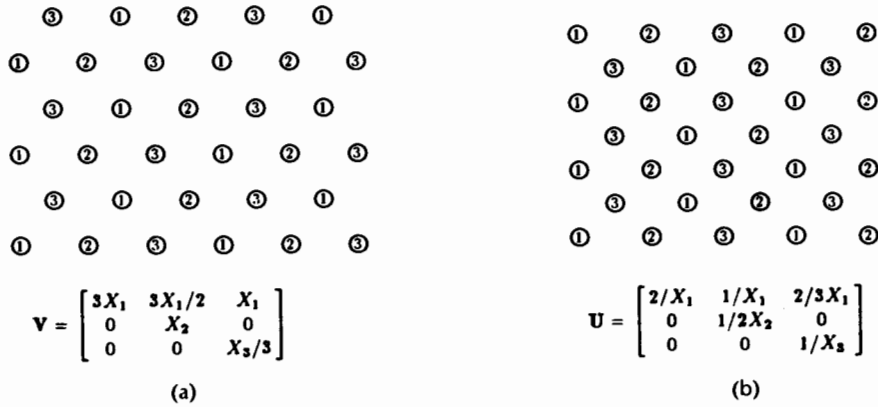
(b)

Fig. 21. (a) Hexagonal lattice with three-field periodicity (HEX3). (b) Reciprocal lattice.

Table 1 Minimum Sampling Density for Alias-Free Sampling of Signals Band-Limited to $|f| < W$

Sampling Structure	Minimum Sampling Density $1/X_1X_2X_3$	X_1	X_2	X_3
ORT	$8W^3 = C$	$1/2W$	$1/2W$	$1/2W$
ALI	$4\sqrt{3}\,W^3 = 0.866C$	$1/2W$	$1/2W$	$1/\sqrt{3}\,W$
BCO	$4\sqrt{2}\,W^3 = 0.707C$	$1/\sqrt{2}\,W$	$1/2\sqrt{2}\,W$	$1/\sqrt{2}\,W$
FCO	$4\sqrt{2}\,W^3 = 0.707C$	$1/2W$	$1/2W$	$1/\sqrt{2}\,W$
QLNS	$6W^3 = 0.75C$	$1/2\sqrt{3}\,W$	$1/2W$	$2/\sqrt{3}\,W$
HEX3	$4\sqrt{2}\,W^3 = 0.707C$	$1/\sqrt{3}\,W$	$1/2W$	$\sqrt{3}/2\sqrt{2}\,W$
QL	$3\sqrt{5}\,W^3 = 0.838C$	$2/\sqrt{15}\,W$	$1/2W$	$1/\sqrt{3}\,W$

cubic (FCC) lattice, which is known to be the densest possible lattice sphere packing. In order to apply these results to the sampling of time-varying imagery, it is necessary to relate measures of temporal frequency to those of spatial frequency. This may not be strictly possible since these are fundamentally different units. An approach suggested by Tonge [28] is to base this relationship on the spatiotemporal response characteristics of the human visual system, and to assume a worst case viewing distance. If the high-frequency portion of the threshold data, as in Fig. 13, is used for this purpose, an approximate equivalence of 1 Hz and 0.62 cycles per degree (or 8.8 c/ph at a viewing
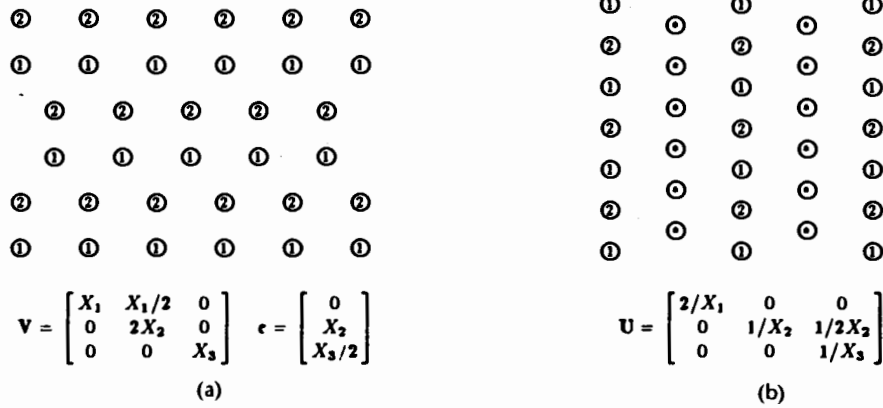
20

$$V = \begin{bmatrix} X_1 & X_1/2 & 0 \\ 0 & 2X_2 & 0 \\ 0 & 0 & X_3 \end{bmatrix} \quad c = \begin{bmatrix} 0 \\ X_2 \\ X_3/2 \end{bmatrix}$$

(a)

$$U = \begin{bmatrix} 2/X_1 & 0 & 0 \\ 0 & 1/X_2 & 1/2X_2 \\ 0 & 0 & 1/X_3 \end{bmatrix}$$

(b)

Fig. 22. (a) Line-quincunx structure (QL). (b) Reciprocal structure. * indicates all multiples of $1/X_3$.
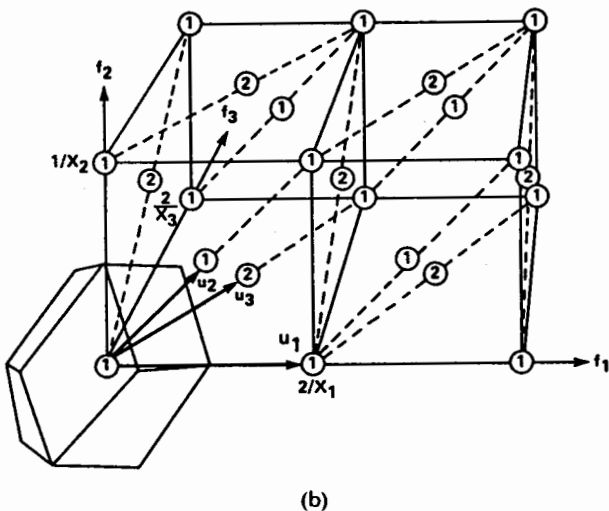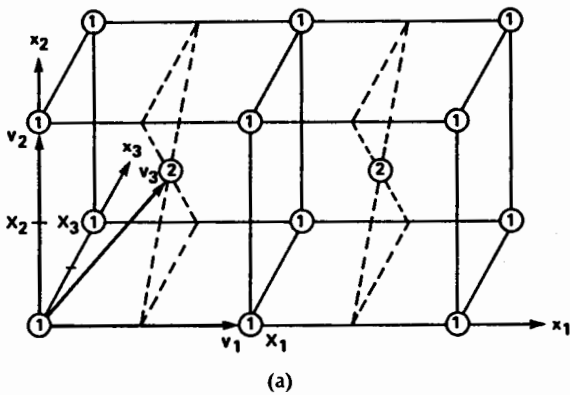

(a)

(b)

Fig. 23. (a) Perspective view of BCO lattice. (b) Perspective view of reciprocal lattice (FCO), showing Voronoi region.

distance of four picture heights) is obtained. This ratio would equate a temporal frequency of 60 Hz with a spatial frequency of 528 c/ph, very close to the ratio used in the NTSC system.

An interesting parameter associated with each sampling structure is the ratio of the spatial sampling density per field to the total spatial sampling density, as seen in the spatial projection of the structure. We refer to this as the interlace

order. This was the basic parameter of interest in the original dot-interlace schemes, and is of particular interest in motion-adaptive interpolation, to be discussed in the next section. For a given spatiotemporal sampling density, the potential spatial resolution in stationary areas increases with interlace order. For the sampling structures of Figs. 16–22, the interlace order is equal to the highest sample index on the diagram of part (a) of the corresponding figure. The HEX3 sampling structure has the highest interlace order (three) among the optimal sphere packing lattices we have given. The HEX4 lattice gives an interlace order of four, with only a small decrease in packing efficiency, making it a potentially attractive structure. However, it remains to be verified that the crawling patterns described in [6] can be eliminated by suitable spatiotemporal interpolation filters.

When sampling a signal which has already been scanned (usually with 2:1 line interlace), the samples must lie on the scanning lines. The structures which are compatible with line-interlaced scanning without field subsampling are ALI, BCO, HEX4, and QL. As the sampling density decreases, it may be advantageous to perform field subsampling to maintain similar resolution in the different dimensions. In this case, the ORT, FCO, and HEX3 patterns may also be used. Tonge [28] has proposed preferred sampling structures (choosing between BCO and FCO) compatible with European scanning standards for different sampling densities.

D. Sampling of Color Signals

Sampling Component Color Signals: The sampling theory which has been presented can be applied in a straightforward manner to sampling the three components of color imagery. The main issues which arise are: which three components should be sampled, and what should be the relative sampling densities? The components which appear to be the most advantageous are the luminance and two chrominance components. From a statistical point of view, most signal energy is contained in the luminance component, and the Y, I, and Q components give similar energy compaction to the Karhunen–Loève transformation [34]. From a perceptual point of view, most successful models of color vision incorporate a luminance channel, and two chrominance channels (for example [35] and [36]). It is well

21

known, that for a given total bandwidth, overall picture quality is maximized by allotting more bandwidth to the luminance than to the chrominance components; this forms the basis for existing color television systems. The recent recommendation for a digital component studio standard [37] allows half the luminance sampling density to each chrominance component (R–Y and B–Y). For many coding systems, a much larger ratio of luminance to chrominance sampling density is used. Further work is required to determine more precisely the optimal allocation of relative spatiotemporal sampling density to the luminance and chrominance components.

Sampling Composite Color Signals: The sampling of composite color video signals presents interesting problems, due to the irregular shape of the spectrum. In composite signals, the luminance and chrominance information is frequency-multiplexed in the three-dimensional frequency space. The form of the three-dimensional spectrum of the PAL signal was derived in [38], and that of the NTSC signal in [5]. Fig. 24 shows an idealized view of the three-
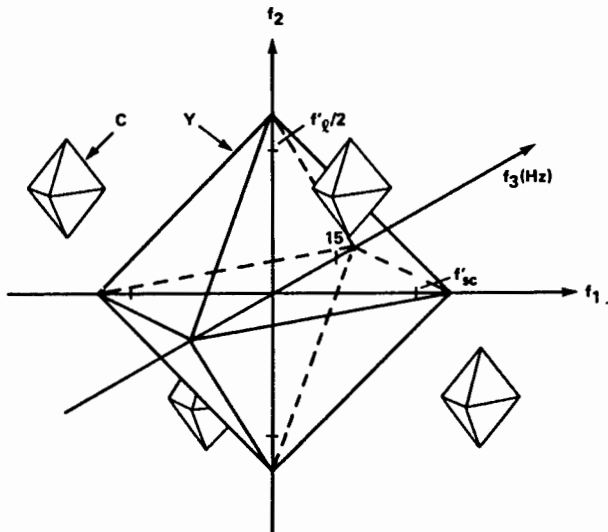

Fig. 24. Three-dimensional spectrum of NTSC signal.

dimensional spectrum of the NTSC signal. The sampling problem is to choose a structure which will optimally pack this spectrum region in three-dimensional frequency space. A number of structures for this purpose, with a horizontal sampling frequency equivalent to $2f_{sc}$, were presented and evaluated in [39]. Although this frequency is "sub-Nyquist" in the one-dimensional sense, appropriate patterns give good separation of the replicated spectra in three dimensions. The preferred sampling pattern was determined to be the body-centered orthorhombic pattern (or field quincunx), and suitable interpolating filters were presented.

IV. FILTERS FOR PREFILTERING AND INTERPOLATION

A. Linear Shift-Invariant Filters

The prefilter and interpolation filters of Fig. 1 are generally hybrid analog/digital filters. The first stage of the prefilter consists of the three-dimensional camera aperture filter, and possibly a one-dimensional analog electrical filter. The output of this stage is sampled on an initial sampling

structure. It may then be digitally filtered, and subsampled to the final sampling structure. At the display, the sampled signal may first be digitally interpolated to a higher sampling density. This is then interpolated to a continuous signal by a one-dimensional analog electrical filter and the display aperture.

There is little flexibility available in shaping the display an camera apertures, whereas there is great flexibility in the design of multidimensional digital filters, and many design techniques are available [7]. Since linear phase is an important requirement in image filtering, finite-impulse response (FIR) filters are most often used, where perfect linear phase can be obtained. Infinite-impulse response (IIR) filters can potentially give equivalent filter performance for a lower order, but introduce problems of phase response and stability. Since the implementation of a filter requires a number of field memories equal to the temporal order of the filter, there may be some advantage to be gained by the use of IIR filters if these problems can be overcome.

There are two basic approaches available for the design of the prefilters and interpolators.

1) Frequency-domain approach: for a given filter order, the impulse response coefficients are chosen so that the frequency response approximates in some way the pass-stop characteristics of the ideal prefilter or interpolator, as described in Section II-C.

2) Spatiotemporal-domain approach: for a given filter order, the prefilter and interpolator coefficients are chosen so that the interpolated signal approximates as closely as possible the original image.

In video applications, the filters are generally of relatively low order, due to speed and memory constraints. The transition regions of the filters are thus not very sharp, so that in-band components are attenuated, while out-of-band components are not completely eliminated. The filters should be optimized on the basis of subjective criteria, in order to give the best tradeoff between resolution loss, aliasing, and ringing. These do not in general carry equal weight in terms of mean-square error. For example, it has been found on the basis of subjective experiments on still pictures that viewers attach more importance to loss of sharpness than to aliasing distortion [40], [41].

Tonge [28] has presented frequency-domain designs for prefilters and interpolators. For the low filter order usually considered, a simple frequency constraint approach is often successful. This can be accomplished by specifying that the frequency response and its derivatives take on specified values at certain frequencies. For example, the response at the origin can be specified to give the desired dc gain, the response at points of the reciprocal lattice in frequency bands to be attenuated can be set to zero, and the derivatives of the response at these frequencies can be set to zero to give maximally flat response in different directions. Each constraint results in a linear equation. If the number of independent constraints equals the number of degrees of freedom of the filter, a unique solution can be found. Otherwise, the remaining coefficients can be varied in some systematic fashion to obtain the best response. This was the approach used in [28]. The McClellan transformation approach has been proposed for the design of two-dimensional prefilters and interpolators, for use with still images [42].

The spatiotemporal approach has been mainly confined to the use of polynomial-type interpolation. Most prominent among these is the cubic B-spline approach [43], which has been used for spatial interpolation. This technique has the advantage of being able to compute interpolated values anywhere, not just on a rationally related lattice. Another spatiotemporal approach is to choose the interpolation coefficients to minimize the mean-square interpolation error, based on a homogeneous random field model for the image signal. This technique has mainly been used for one-dimensional interpolation [44], [45], but is easily extended to three dimensions.

B. Nonlinear Motion-Dependent Filters

The preceding sections on subsampling and interpolation assume that the signal is reconstructed by means of a linear three-dimensional low-pass filter. The specific nature of time-varying imagery can be exploited to obtain improved reconstruction techniques. Specifically, a time-varying scene generally consists of a fixed background and a number of moving objects. The temporal variation in the imagery is caused by the motion of the objects in the scene. This section presents reconstruction techniques which adapt to the motion. Two main classes of algorithms are considered. In the first, termed *motion-adaptive* processing, a motion *detection* operation is performed, and different filtering algorithms are used in moving and stationary parts of the scene. In the second, *motion-compensated processing*, the displacement of the objects from frame to frame is estimated, and this motion *estimation* is used to determine the filtering which is performed.

Motion-Adaptive Interpolation: Motion-adaptive interpolation can be used when it is desired to interpolate to a higher density sampling structure which has the same spatial projection. The maximum interpolation ratio which can be obtained in this way is equal to the interlace order of the original sampling structure. The idea is to perform temporal interpolation in fixed areas, interpolating a missing sample from samples at the same spatial location in previous and subsequent fields. In moving areas, where this would not give good results, a spatial interpolator is used. The basic concept is shown in Fig. 25. A motion detector generates a motion index M, largely based on the interframe difference in a neighborhood of the sample to be interpolated. To avoid sudden switching between the two

types of interpolators, a weighted sum of the two, with weights determined by the motion index, is used. Thus if $\hat{u}_t(x)$ is a temporal interpolation and $\hat{u}_s(x)$ is a spatial interpolation at point x, the motion-adaptive interpolation is given by

$$\hat{u}(x) = \alpha(M)\hat{u}_s(x) + (1 - \alpha(M))\hat{u}_t(x). \tag{63}$$

Fig. 26 shows the general form of the function $\alpha(M)$ which relates the weighting factor to the motion index. If the motion index is small, α is close to zero, and the interpola-
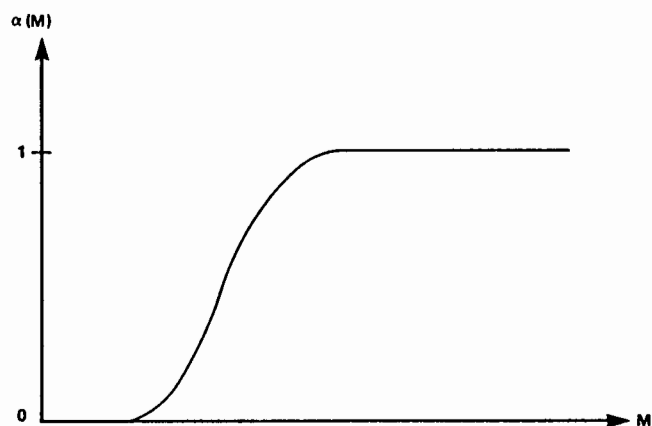

Fig. 26. Nonlinear function of motion index for motion-adaptive interpolation.

tion is mainly temporal. As the motion index increases, α approaches unity, and the interpolation becomes mainly spatial. This type of interpolation can give significantly better results than fixed interpolation, especially in stationary areas where the eye is most critical. The performance of motion-adaptive interpolation is critically dependent on the motion detection algorithm, and most research on this technique centers on this aspect. Motion-adaptive interpolation has been applied in the interpolation from the HEX4 structure to the orthorhombic structure which contains it, in the subsampling subsystem of a low-rate coder for video conferencing [46]. It has also been applied in the interpolation from interlaced scanning to sequential scanning with twice the number of lines per field, for improved display performance [47]–[49]. Motion-adaptive processing, based on the same principles, has also been applied to noise reduction [50] and luminance/chrominance demultiplexing in composite signals [49].

Motion-Compensated Interpolation: Motion compensation is a technique which has received considerable attention recently for application in predictive coding of time-varying imagery (e.g., [51]–[53]), and in noise reduction [54]. This technique can be used to improve the performance of interpolation in moving areas of the picture, especially when pure temporal subsampling has been used, and the spatial interpolation part of the motion-adaptive interpolator is not relevant. The basis of the technique is to estimate the motion of the objects in the scene, and to interpolate a missing field using the corresponding object points in the previous and subsequent transmitted fields. The concept was described by Gabor and Hill [55], using interpolation only along a given scan. The application to general two-dimensional motion was described in a series of papers at
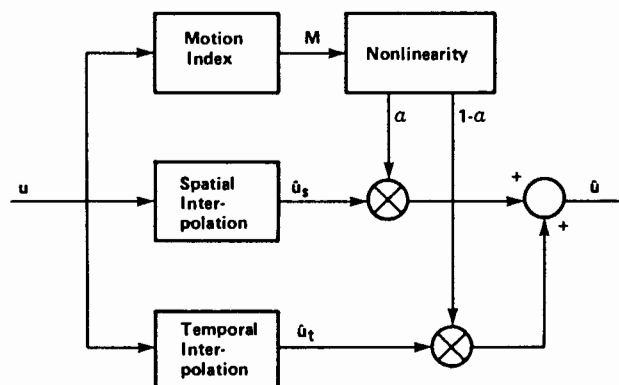

Fig. 25. Motion-adaptive interpolation system.

the 1981 Picture Coding Symposium [56]–[58]. Specific algorithms to perform motion-compensated interpolation are disclosed in [59], and applications in low bit-rate coding are given in [60], [61].

With conventional temporal interpolation, a missing field is reconstructed by forming a weighted combination of pels at the same spatial location in previous and subsequent transmitted fields (using spatial interpolation if necessary). When the subsampling factor is large, this results in both jerkiness (a result of aliasing) and blurring. Motion-compensated interpolation can significantly reduce these effects. The general approach is illustrated in Fig. 27 for the case of 2:1 field subsampling. For each point in the output sampling structure of the nontransmitted field, such as pel *A*, the spatial location of the corresponding point in the
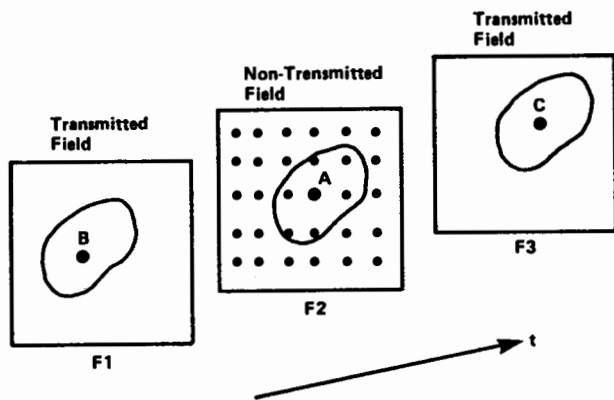

Fig. 27. Principle of motion-compensated interpolation. Point *A* in field *F2* is interpolated using point *B* in field *F1* and point *C* in field *F3*, after having estimated the displacement vector.

previous and subsequent transmitted fields must be estimated, *B* and *C* in Fig. 27. It is assumed that the points *BAC* lie on a straight line. A linear combination of the intensities of the points at *B* and *C* is used to interpolate the value at *A*. Many techniques have been developed for displacement estimation, the most important ones being matching [62], [63], the method of spatiotemporal gradients [64], [65], and steepest descent recursive estimation [51]. A review of displacement estimation can be found in [66]. Any of these methods can be applied to motion-compensated interpolation; a matching approach is used in [60], while the steepest descent approach has been used in [59] and [61].

It is important to recognize that there is no correction for errors made in the displacement estimation, as there is in predictive coding and noise reduction applications, and so it is imperative that the displacement field used be as accurate as possible, or else artifacts will appear in the image. Another observation is that where a signal may be considered undersampled and aliased from the point of view of a homogeneous random field, it can be accurately reconstructed on the basis of a more complex model.

V. CONCLUSION

This paper has presented an overview of techniques for sampling and reconstruction of time-varying imagery. A feature of the presentation is the use of lattices to describe

the sampling process, greatly facilitating the subsequent analysis of digital signal processing algorithms. A good understanding of the principles involved will be required to design systems capable of providing the highest image quality possible for a given transmission format. This is of great importance at this time with the advent of enhanced-quality and high-definition television system proposals, discussed elsewhere in this issue. Further work is required to arrive at three-dimensional linear and nonlinear filters capable of giving the desired image quality with reasonable hardware complexity.

REFERENCES

[1] J. W. S. Cassels, *An Introduction to the Geometry of Numbers.* Berlin, Germany: Springer-Verlag, 1959.
[2] J. M. Ziman, *Principles of the Theory of Solids.* Cambridge, UK: Cambridge Univ. Press, 1972.
[3] D. P. Petersen and D. Middleton, "Sampling and reconstruction of wave-number-limited functions in *N*-dimensional Euclidean spaces," *Informat. Contr.*, vol. 5, pp. 279–323, 1962.
[4] N. T. Gaarder, "A note on the multidimensional sampling theorem," *Proc. IEEE*, vol. 60, pp. 247–248, Feb. 1972.
[5] E. Dubois, M. S. Sabri, and J.-Y. Ouellet, "Three-dimensional spectrum and processing of digital NTSC color signals," *SMPTE J.*, vol. 91, pp. 372–378, Apr. 1982.
[6] F. Kretz and J. Sabatier, "Échantillonnage des images de télévision: analyse dans le domaine spatio-temporel et dans le domaine de Fourier," *Ann. Télécommunications*, vol. 36, pp. 231–273, Mar.–Apr. 1981.
[7] D. E. Dudgeon and R. M. Mersereau, *Multidimensional Digital Signal Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1984.
[8] J. P. Allebach, "Design of antialiasing patterns for time-sequential sampling of spatiotemporal signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, pp. 137–144, Feb. 1984.
[9] W. Rudin, *Fourier Analysis on Groups.* New York: Interscience, 1962.
[10] L. E. Franks, *Signal Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1969.
[11] R. M. Mersereau and T. C. Speake, "The processing of periodically sampled multidimensional signals," *IEEE Trans. Acoust., Speech, Signal Processing.*, vol. ASSP-31, pp. 188–194, Feb. 1983.
[12] A. Papoulis, *Signal Analysis.* New York: McGraw-Hill, 1977.
[13] B. Wendland, "Extended definition television with high picture quality," *SMPTE J.*, vol. 92, pp. 1028–1035, Oct. 1983.
[14] R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1983.
[15] P. Mertz and F. Gray, "A theory of scanning and its relation to the characteristics of the transmitted signal in telephotography and television," *Bell Syst. Tech. J.*, vol. 13, pp. 464–515, 1934.
[16] A. Macovski, "Spatial and temporal analysis of scanned systems," *Appl. Opt.*, vol. 9, pp. 1906–1910, Aug. 1970.
[17] A. H. Robinson, "Multidimensional Fourier transforms and image processing with finite scanning apertures," *Appl. Opt.*, vol. 12, pp. 2344–2352, Oct. 1973.
[18] D. E. Pearson, *Transmission and Display of Pictorial Information.* New-York: Wiley, 1975.
[19] Y. Guinet, "Propriétés spectrales et performance du système télévisuel," *Radiodiffusiontélévision*, "1ère partie: Modélisation des systèmes télévisuels scalaires à support continu," no.

56, pp. 9–14, 1979; "2e partie: Propriétés générales des balayages à entrelacement," no. 57, pp. 49–57, 1979; "3e partie: Systèmes scalaires à support continu et à entrelacement de trame d'ordre deux," no. 59, pp. 30–40, 1979; "4ème partie: Modulation du point mobile de balayage par un signal périodique. 1. Le cas du signal sinusoidal," no. 66, pp. 47–55, 1981.

[20] L. D. Miller, "A new method of specifying the resolving power of television camera tubes using the RCA P-300 test chart," SMPTE J., vol. 89, pp. 249–256, Apr. 1980.

[21] O. H. Schade, Sr., "Image reproduction by a line raster process," in Perception of Displayed Information, L. M. Biberman, Ed. New York: Plenum, 1973, pp. 233–278.

[22] G. A. Reitmeier, "The effects of analog filtering on the picture quality of component digital television systems," SMPTE J., vol. 90, pp. 949–955, Oct. 1981.

[23] D. H. Kelly, "Motion and vision. II. Stabilized spatio-temporal threshold surface," J. Opt. Soc. Amer., vol. 69, pp. 1340–1349, Oct. 1979.

[24] F. X. J. Lukas and Z. L. Budrikis, "Picture quality prediction based on a visual model," IEEE Trans. Commun., vol. COM-30, pp. 1679–1692, July 1982.

[25] N. A. Diakides, "Phosphor screens," in Electronics Engineers' Handbook, D. G. Fink, Ed. New York: McGraw Hill, pp. 11-33–11-39, 1975.

[26] T. Mitsuhashi, "A study of the relationship between scanning specifications and picture quality," NHK, NHK Laboratories Note, no. 256, Oct. 1980.

[27] G. J. Tonge, "The television scanning process," SMPTE J., vol. 93, pp. 657–666, July 1984.

[28] ———, "Three-dimensional filters for television sampling," Independent Broadcasting Authority, Experimental and Development Rep. 117/82, June 1982.

[29] E. F. Brown, "Low-resolution TV: Subjective comparison of interlaced and noninterlaced pictures," Bell Syst. Tech. J., vol. 46, pp. 199–232, Jan. 1967.

[30] G. G. Gouriet, "Dot-interlaced television," Electron. Eng., vol. 24, pp. 166–171, Apr. 1952.

[31] R. C. Brainard, F. W. Mounts, and B. Prasada, "Low-resolution TV: Subjective effects of frame repetition and picture replenishment," Bell Syst. Tech. J., vol. 46, pp. 261–271, Jan. 1967.

[32] J. Sabatier and F. Kretz, "Sampling the components of 625-line colour-television signals," E. B. U. Rev.—Tech., no. 171, pp. 212–225, Oct. 1978.

[33] G. J. Tonge, "The sampling of television images," Independent Broadcasting Authority, Experimental and Development Rep. 112/81, May 1981.

[34] W. K. Pratt, "Spatial transform coding of color images," IEEE Trans. Commun. Technol., vol. COM-19, pp. 980–992, Dec. 1971.

[35] W. Frei and B. Baxter, "Rate-distortion coding simulation for color images," IEEE Trans. Commun., vol. COM-25, pp. 1385–1392, Nov. 1977.

[36] O. D. Faugeras, "Digital color image processing within the framework of a human visual model," IEEE Trans. Acoust., Speech, Signal Process., vol. ASSP-27, pp. 380–393, Aug. 1979.

[37] F. Davidoff, "Digital television coding standards," Proc. Inst. Elec. Eng., pt. A, vol. 129, pp. 403–412, Sept. 1982.

[38] J. O. Drewery, "The filtering of luminance and chrominance signals to avoid cross-colour in a PAL colour system," BBC Eng., pp. 8–39, Sept. 1976.

[39] J.-Y. Ouellet and E. Dubois, "Sampling and reconstruction of NTSC video signals at twice the color subcarrier frequency," IEEE Trans. Commun., vol. COM-29, pp. 1823–1832, Dec. 1981.

[40] J. N. Ratzel, "The discrete representation of spatially continuous images," Massachusetts Institute of Technology, Ph.D. dissertation, Aug. 1980.

[41] R. J. Arguello, H. B. Kessler, and H. R. Sellner, "The effect of sampling, optical transfer function shape, and anisotropy on subjective image quality," Opt. Eng., vol. 21, pp. 23–29, Jan./Feb. 1982.

[42] T. C. Chen and R. J. P. de Figueiredo, "Image decimation and interpolation techniques based on frequency domain analysis," IEEE Trans. Commun., vol. COM-32, pp. 479–484, Apr. 1984.

[43] H. S. Hou and H. C. Andrews, "Cubic splines for image interpolation and digital filtering," IEEE Trans. Acoust., Speech, Signal Process., vol. ASSP-26, pp. 508–517, Dec. 1978.

[44] G. Oetken, T. W. Parks, and H. W. Schüssler, "New results in the design of digital interpolators," IEEE Trans. Acoust., Speech, Signal Process., vol. ASSP-23, pp. 301–309, June 1975.

[45] A. D. Polydoros and E. N. Protonotarios, "Digital interpolation of stochastic signals," IEEE Trans. Circuits Syst., vol. CAS-26, pp. 916–922, Nov. 1979.

[46] K. Takikawa, "Simplified 6.3 Mbit/s codec for video conferencing," IEEE Trans. Commun., vol. COM-29, pp. 1877–1882, Dec. 1981.

[47] C. P. Sandbank and M. E. B. Moffat, "High-definition television and compatibility with existing standards," SMPTE J., vol. 92, pp. 552–561, May 1983.

[48] G. Tonge, "Signal processing for higher-definition television," IBA Tech. Rev., vol. 21, pp. 13–26, Nov. 1983.

[49] M. Achiha, K. Ishikura, and T. Fukinuki, "A motion-adaptive high-definition converter for NTSC color TV signals," SMPTE J., vol. 93, pp. 470–476, May 1984.

[50] J. R. Sanders, "Fully adaptive noise reduction for a television network," Television: J. Royal Television Soc., vol. 18, pp. 29–33, May/June 1980.

[51] A. N. Netravali and J. D. Robbins, "Motion-compensated television coding: Part I," Bell Syst. Tech. J., vol. 58, pp. 631–670, Mar. 1979.

[52] S. Sabri, "Movement compensated interframe prediction for NTSC color TV signals," IEEE Trans. Commun., vol. COM-32, pp. 954–968, Aug. 1984.

[53] T. Koga, A. Hirano, K. Iinuma, Y. Iijima, and T. Ishiguro, "A 1.5 Mb/s interframe codec with motion compensation," in Rec. Int. Conf. Commun., pp. D8.7.1–D8.7.5, June 1983.

[54] E. Dubois and S. Sabri, "Noise reduction in image sequences using motion-compensated temporal filtering," IEEE Trans. Commun., vol. COM-32, pp. 826–831, July 1984.

[55] D. Gabor and P. C. J. Hill, "Television band compression by contour interpolation," Proc. Inst. Elec. Eng., vol. 108, pt. B, pp. 303–315, May 1961.

[56] A. N. Netravali and J. D. Robbins, "Motion adaptive interpolation of television frames," in Rec. Picture Coding Symp., p. 115, June 1981.

[57] H. C. Bergmann, "Motion-adaptive interpolation of eliminated TV fields," in Rec. Picture Coding Symp., pp. 116–117, June 1981.

[58] R. Lippmann, "Air-to-ground TV at low frame rates extended by interpolative frame regeneration," in Rec. Picture Coding Symp., pp. 120–121, June 1981.

[59] A. N. Netravali and J. D. Robbins, "Video signal interpolation using motion estimation," U.S. Patent 4 383 272, May 10, 1983.

[60] A. Furukawa, T. Koga, and K. Iinuma, "Motion-adaptive interpolation for videoconference pictures," in Proc. Int. Conf. Commun., pp. 707–710, May 1984.

[61] S. Sabri, K. Cuffling, and B. Prasada, "Coding of video signals at 50 kb/s using motion compensation techniques," in Proc. IEEE Military Comm. Conf., pp. 809–816, Nov. 1983.

[62] J. R. Jain and A. K. Jain, "Displacement measurement and its application in interframe image coding," IEEE Trans. Commun., vol. COM-29, pp. 1799–1808, Dec. 1981.

[63] Y. Ninomiya and Y. Ohtsuka, "A motion-compensated interframe coding scheme for television pictures," IEEE Trans. Commun., vol. COM-30, pp. 201–211, Jan. 1982.

[64] C. Cafforio and F. Rocca, "Methods for measuring small displacements of television images," IEEE Trans. Informat. Theory, vol. IT-22, pp. 573–579, Sept. 1976.

[65] R. Paquin and E. Dubois, "A spatio-temporal gradient method for estimating the displacement field in time-varying imagery," Comput. Vision, Graph. Image Process., vol. 21, pp. 205–221, 1983.

[66] W. B. Thompson and S. T. Barnard, "Lower-level estimation and interpretation of visual motion," Computer, vol. 14, pp. 20–28, Aug. 1981.