

Analysis of frame-compatible subsampling structures for efficient 3DTV broadcast

Raúl Mohedano · Carlos Muñoz · Julián Cabrera ·
Fernando Jaureguizar · Narciso García

Abstract The evolution of the television market is led by 3DTV technology, and this tendency can accelerate during the next years according to expert forecasts. However, 3DTV delivery by broadcast networks is not currently developed enough, and acts as a bottleneck for the complete deployment of the technology. Thus, increasing interest is dedicated to stereo 3DTV formats compatible with current HDTV video equipment and infrastructure, as they may greatly encourage 3D acceptance. In this paper, different subsampling schemes for HDTV compatible transmission of both progressive and interlaced stereo 3DTV are studied and compared. The frequency characteristics and preserved frequency content of each scheme are analyzed, and a simple interpolation filter is specially designed. Finally, the advantages and disadvantages of the different schemes and filters are evaluated through quality testing on several progressive and interlaced video sequences.

Keywords 3DTV · Frame-compatible format · Side-by-side · Subsampling · Interpolation

1 Introduction

More and more, 3D video technology is becoming an important part of our daily life: movie theaters are offering a growing variety of 3D films; new generation video game companies are starting using 3D for improved immersive experience (even a handheld console including

R. Mohedano (✉) · C. Muñoz · J. Cabrera · F. Jaureguizar · N. García
Grupo de Tratamiento de Imágenes, Universidad Politécnica de Madrid, E.T.S.I. Telecomunicación,
28040 Madrid, Spain
e-mail: rmp@gti.ssr.upm.es

C. Muñoz
e-mail: cmp@gti.ssr.upm.es

J. Cabrera
e-mail: julian.cabrera@gti.ssr.upm.es

F. Jaureguizar
e-mail: fjn@gti.ssr.upm.es

N. García
e-mail: narciso@gti.ssr.upm.es

a specific 3D viewing display has been recently released Nintendo® 2011); and, of course, all major TV set manufacturers are dedicating increasing resources to position themselves as leaders of the 3D consumer market, which, according to expert forecasts (King 2011), will grow exponentially during the next years.

More specifically, the 3DTV system with faster adoption has clearly been the stereo format, due mainly to its simplicity in terms of content acquisition, processing and also display. In fact, this is the technology currently used in commercially available TV sets via simple and affordable stereoscopic displays (i.e. using active or passive glasses to filter the corresponding view Merkle et al. 2010). In HDTV stereo video, each channel is acquired at the common resolution and frame-rate for 2D HDTV formats (e.g. 1080p, 1080i, 720p), which would roughly double the bandwidth requirements of an equivalent 2D HDTV video signal. Specific coding formats considering redundancy between frames to alleviate these excessive bandwidth requirements have been (and are yet being) developed. However, their practical adoption is far from achieved, as they require specific equipment and transmission signaling, incompatible specially with actual systems broadcasting systems.

Fortunately, another essential characteristic that makes stereo so interesting is that it allows the definition of coding and transmission formats totally compatible with current 2DTV equipment and technology (ITU-R Broadcasting Service 2009): for instance, decoders based on HDMI version 1.3 designed originally for single-view TV but widely available nowadays. These formats are usually referred to in the literature as *frame-compatible* (Vetro et al. 2011). This compatibility encourages greatly the penetration of stereo-based 3DTV in the consumer electronics market, and therefore recent interfaces and video standards such as HDMI 1.4 (HDMI Licensing, LLC. 2009) and H.264 (ITU-T and ISO/IEC JTC1 2010) include specific support for describing different allowed options and arrangements.

All frame-compatible formats combine both left and right channels into one single 2D-like video stream whose resolution and frame-rate equals that of each original channel: in that way, 3D stereo streams can be multiplexed, transmitted and processed through conventional equipment and delivery channels for 2D video, and finally demultiplexed again. Multiplexing can be performed either temporally, by alternately suppressing one frame of each channel, or spatially, by halving the resolution of both channels at every time step. Apparently, spatial multiplexing has been clearly preferred, and among all the possible spatial arrangements, *Side-by-Side* (SbS, horizontal resolution halving) and *Top-and-Bottom* (TaB, vertical halving, analogous to SbS), have been the most frequent options (ITU-R Broadcasting Service 2009). However, these are only two of the multiple spatial multiplexing arrangements that can be proposed for frame packing.

Here, we propose and compare several spatial packing schemes for both progressive and interlaced stereo 3DTV. A detailed analysis of the different discussed arrangements is performed in both spatial-temporal and frequency domains, placing special emphasis on the spectral limitations and capabilities of each one through the study of its associated Voronoi cell (Dubois 1985). Stemming from it, and imposing purely-spatial processing for maximum simplicity in terms of computational cost, we also design specific intraframe filters (intrafield for interleaved schemes) fulfilling the requirements of each structure for frequency preservation, which prove themselves effective even for small filter sizes thanks to a newly proposed normalization step. This design is first conducted and meticulously presented for progressive video schemes, and later extended to the proposed interlaced structures, discussing the similarities and differences between them. Finally, the proposed packing schemes along with the proposed filters are tested on different reference sequences, and the obtained results are interpreted to obtain valuable conclusions on the use of subsampling schemes and interpolation filters.

The remainder of this paper is organized as follows: Sect. 2 presents and analyzes different frame-compatible spatial packing schemes for progressive 3D video, while Sect. 3 preforms the analogous analysis for several schemes for interlaced 3D stereo. Design of anti-aliasing and interpolating filters for the previously discussed schemes is addressed in Sect. 4. The performance of the different schemes and filters is discussed in Sect. 5, and the main final conclusions of this work are included in Sect. 6.

Notation: We will represent lattices and sublattices throughout the paper using the notation presented by Dubois (1985) in his seminal paper: each lattice Λ will be described by means of a matrix V whose columns contain the coordinates of the vectors of its generating basis $B = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ with respect to the canonical basis for \mathbb{R}^3 ; whereas, the corresponding reciprocal lattice could be analogously represented by a matrix U such that $U^T V = I$, where I represents the identity matrix. Additionally, the two first components of the original domain, x and y , represent respectively the horizontal and vertical directions, and the third component, t , represents the temporal evolution; the corresponding transformed coordinates will be represented, respectively, as f_x , f_y and f_t .

Relationship between spatial and temporal dimensions: The frequency characteristics of the subsampling schemes described throughout this paper depend on the relationship between the three dimensions involved: horizontal (x) and vertical (y) spatial directions, and time (t). Here, the criteria followed to define that relative scaling between dimensions are stated and justified.

In principle, the relationship between the two spatial dimensions x and y can be straightforwardly established. As suggested in Sullivan et al. (1991), although the human visual system cannot be considered isotropic, it does present similar perceptual sensitivity and limits for both horizontal and vertical directions. Therefore, it seems safe to assume a 1:1 relationship between x and y . Additionally, current acquisition and display video devices consider square pixels: thus, subsampling lattice definitions throughout the paper will assume equal vertical and horizontal base units $X = Y$.

Unfortunately, establishing a relationship between spatial and temporal coordinates is by no means an obvious task, as time and space are dimensions of an extremely different nature. Although any possible criterium for inferring their relationship could be arguable, we consider the spatial-temporal behavior of the human visual system to be a good model for all our considerations. The frequency response of the human eye is expressed in the specialized literature by means of contrast sensitivity functions (CSF) expressing the minimum distinguishable contrast at each spatial and/or temporal frequency measured, respectively, in *cycles per degree* (cpd) and *hertz* (Hz). These functions are obtained (or checked) empirically by subjective testing (Kelly 1979, 1983), and show a bandpass shape in both spatial and temporal dimensions.

The CSF varies slightly depending on the viewing distance (Mannos and Sakrison 1974) due to optical effects. Additionally, and partly as a consequence of that dependence, it has been reported that certain resolution-to-viewing distances are more comfortable for the user (ITU-R Broadcasting Service 2002). However, it seems safe enough to study human visual response at average conditions. In that direction, Fairchild (2005) studied independently spatial and temporal contrast sensitivity. According to his results, chromatic contrast sensitivity decays faster than luminance sensitivity for both spatial and temporal variations, so the latter seems to mark the limits of human vision. Additionally, the cut-off frequency for the spatial CSF for luminance is about 60 cpd, whereas this limit is about 60 Hz for the corresponding temporal CSF (Díez-Ajenjo and Capilla 2010). It is maybe more interesting to consider the frequency where the CSF decays to 10% of its maximum value: that decay is produced, in both spatial

and temporal responses, at 50–60% of the cut-off frequency. Thus, both perceptual limits suggest a 1:1 relation between spatial frequency units measured in cpd and temporal frequencies measured in Hz.

Therefore, for usual viewing conditions, we can assume that temporal and spatial base units are equal (measured, respectively, in cpd and Hz), so $X = Y = T$. In addition, for simplicity but with any loss of generality, we can consider normalized dimensions ($X = Y = T = 1$) whenever it is required in this paper.

2 Progressive 3D video format schemes

As for frame-compatible for progressive 3D video, we will consider three different sublattices of the basic orthorhombic lattice Λ_O (whose density doubles that of the sublattices) for study. We will respectively denote these schemes as: horizontal Λ_{pH} (horizontally-aligned sampling of 2:1), stationary quincunx Λ_{pQ} , and non-stationary quincunx Λ_{pN} (referred to as FCO, Face-Centered Orthorhombic lattice, in Dubois 1985). The considered sublattices are described by the matrices

$$\mathbf{V}_{pH} = \begin{pmatrix} 2X & 0 & 0 \\ 0 & Y & 0 \\ 0 & 0 & T \end{pmatrix}, \quad \mathbf{V}_{pQ} = \begin{pmatrix} 2X & X & 0 \\ 0 & Y & 0 \\ 0 & 0 & T \end{pmatrix}, \quad \mathbf{V}_{pN} = \begin{pmatrix} 2X & X & X \\ 0 & Y & 0 \\ 0 & 0 & T \end{pmatrix}, \quad (1)$$

where X and Y represent, respectively, the separation of the orthorhombic lattice in the horizontal and vertical axes, and $1/T$ is the frame-rate of the progressive video. The spatial-temporal layout of these subsampling lattices is shown in Fig. 1. Their frequency layout is shown, respectively, in Figs. 2, 3 and 4, where each corresponding Voronoi cell is marked as a solid polyhedron.

The three considered subsampling lattices have the same overall density (half as that of Λ_O): however, they present different frequency characteristics. In principle, according to Dubois (1985), the spectral preservation capabilities of a lattice are directly related to its ability to pack non-overlapping spheres centered at the points of reciprocal lattice, which is directly related to the corresponding Voronoi cell of corresponding reciprocal lattice: therefore, the non-stationary quincunx sublattice seems to show the best characteristics of all in terms of frequency preservation, followed by the stationary quincunx (see Figs. 2, 3, 4). However, multiple perceptual experiments suggest that the human visual system is much

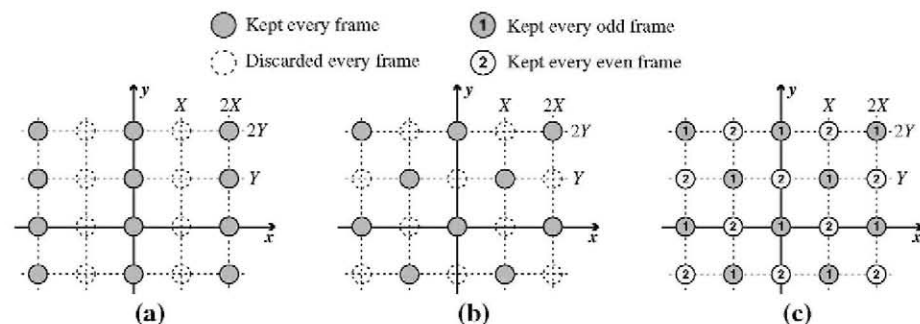


Fig. 1 Spatial-temporal layout of the lattices: **a** horizontal Λ_{pH} , **b** stationary quincunx Λ_{pQ} and **c** non-stationary quincunx Λ_{pN} for progressive video subsampling

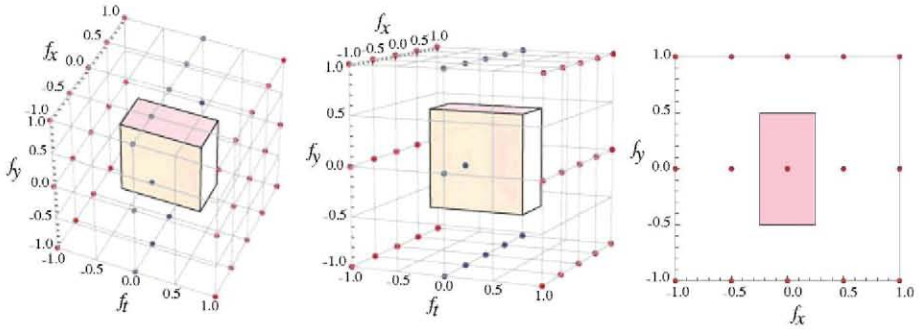


Fig. 2 Frequency layout of the horizontal lattice Λ_{pH} for progressive video subsampling, with its corresponding Voronoi cell depicted as a *solid* polyhedron, from three different views

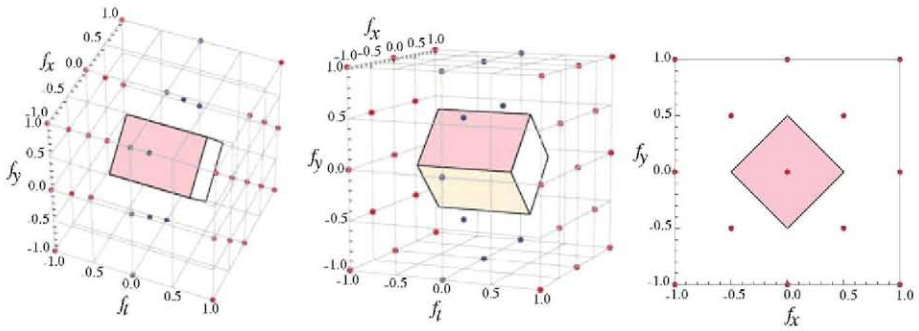


Fig. 3 Frequency layout of the stationary quincunx lattice Λ_{pQ} for progressive video subsampling, with its corresponding Voronoi cell depicted as a *solid* polyhedron, from three different views

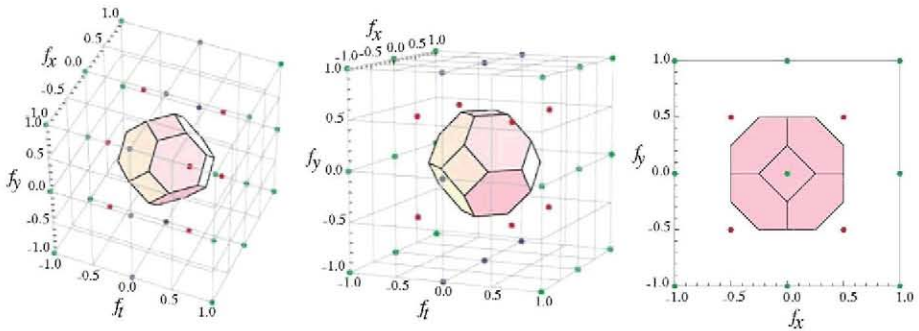


Fig. 4 Frequency layout of the non-stationary quincunx lattice Λ_{pN} for progressive video subsampling, with its corresponding Voronoi cell depicted as a *solid* polyhedron, from three different views

more sensitive to purely horizontal and vertical spatial frequencies than to diagonal variations (Sullivan et al. 1991): thus, the subjective difference between the non-stationary and the stationary quincunx sublattices could be lower than initially expected, as suggested in the $f_t = 0$ central sections of Fig. 5b, c.

Low-complexity filters are also strongly desirable to ensure a simple integration of frame-compatible stereo formats with existing 2DTV equipment: thus, purely-spatial

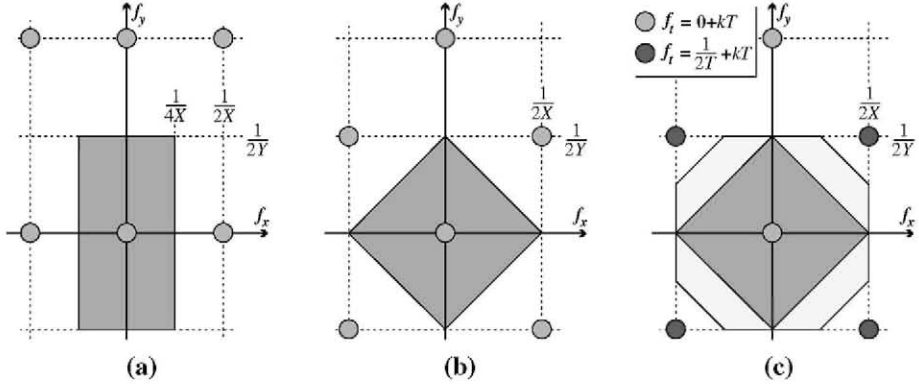


Fig. 5 Purely spatial frequencies of the progressive subsampling lattices: **a** horizontal Λ_{pH} , and **b** stationary Λ_{pQ} and **c** non-stationary quincunx Λ_{pN} . *Shadowed*: cut of the 3D Voronoi region at $f_t=0$. *Dark-grey*: preserved frequencies for intra-frame-only processing

filters processing each frame independently are specially interesting. Imposing that condition, only filters whose spectral response shows translational symmetry along the f_t axis can be achieved. Thus, the optimal Voronoi cells (in terms of sphere packing) of each sublattice depicted in Figs. 2, 3 and 4 can only be theoretically reached for the Λ_{pH} and Λ_{pQ} sublattices. As for the non-stationary quincunx lattice Λ_{pN} , the desirable frequency region given intra-frame filters will coincide with the Voronoi cell of the stationary quincunx Λ_{pQ} (see Fig. 5b, c). Therefore, from now on, we will address jointly both quincunx lattices Λ_{pH} and Λ_{pQ} , and we will not explicitly include the time-related dimensions t and f_t .

3 Interlaced 3D video format schemes

As for interlaced video frame-compatible format is concerned, we will consider three different subsampling structures of the basic interlaced lattice Λ_I , which we will denote as: horizontal Λ_{iH} , BCO (Body-Centered Orthorhombic) quincunx Λ_{iB} and LQ (Line-Quincunx) Λ_{iL} . Using the same notation as in Sect. 2 for lattices, the two first schemes can be described by the matrices

$$\mathbf{V}_{iH} = \begin{pmatrix} 2X & 0 & 0 \\ 0 & 2Y & Y \\ 0 & 0 & \frac{T}{2} \end{pmatrix}, \quad \mathbf{V}_{iB} = \begin{pmatrix} 2X & 0 & X \\ 0 & 2Y & Y \\ 0 & 0 & \frac{T}{2} \end{pmatrix}. \quad (2)$$

However, the last scheme, Λ_{iL} , is not properly a lattice, but the combination of two cosets instead (Dubois 1985). Thus, it is described by means of its basis and the displacement \mathbf{c} of the second coset, and then

$$\mathbf{V}_{iL} = \begin{pmatrix} 2X & X & 0 \\ 0 & 2Y & 0 \\ 0 & 0 & T \end{pmatrix}, \quad \mathbf{c} = \left[0, Y, \frac{T}{2} \right]. \quad (3)$$

The spatial-temporal layout of these subsampling schemes is shown in Fig. 6. Their frequency layout is displayed, respectively, in Figs. 7, 8 and 9, with their Voronoi cells depicted as a solid polyhedron.

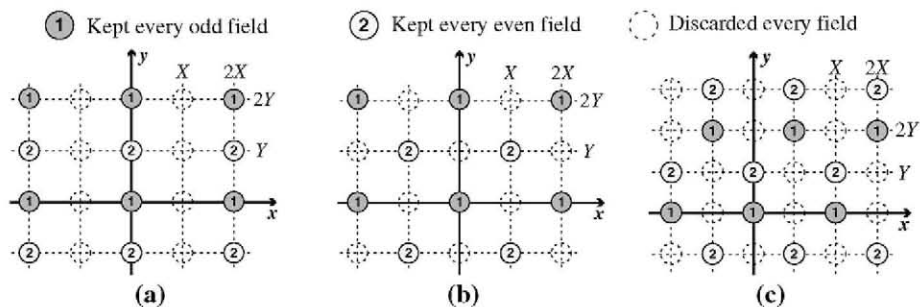


Fig. 6 Spatial-temporal layout of the structures: **a** horizontal Λ_{iH} , **b** BCO quincunx Λ_{iB} and **c** e.L.Q. quincunx Λ_{iL} for interlaced video subsampling

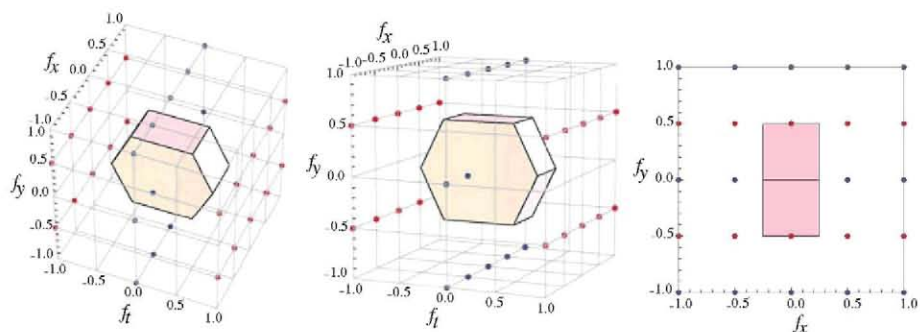


Fig. 7 Frequency layout of the horizontal lattice Λ_{iH} for interlaced video subsampling, with its corresponding Voronoi cell depicted as a solid polyhedron, from three different views

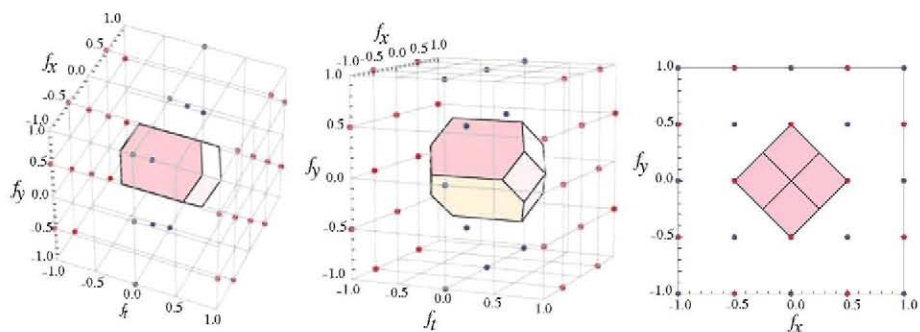


Fig. 8 Frequency layout of the BCO quincunx lattice Λ_{iB} for interlaced video subsampling, with its corresponding Voronoi cell depicted as a solid polyhedron, from three different views

We will assume again that the frequency preservation capabilities of a certain lattice are directly related to the Voronoi cell of its corresponding reciprocal structure. Then, a careful analysis of the three-dimensional Voronoi regions for the three described structures (Figs. 7, 8, 9) indicates that, ideally, the BCO lattice can preserve a greater range of spatial frequencies than the other two considered structures, keeping specially pure horizontal frequencies (see shadowed areas in Fig. 10a, b, c). However, as discussed in Sect. 2 for progressive video subsampling schemes, it is desirable in terms of filtering complexity to restrict the analysis to

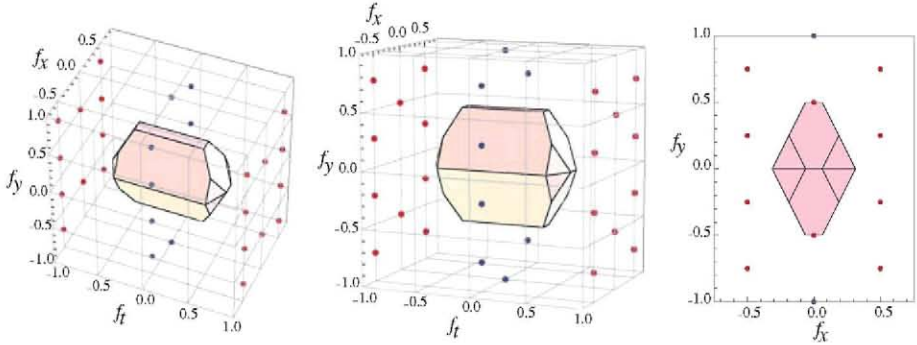


Fig. 9 Frequency layout of the I.Q quincunx scheme Λ_{iL} for intelaced video subsampling, with its corresponding Voronoi cell depicted as a solid polyhedron, from three different views

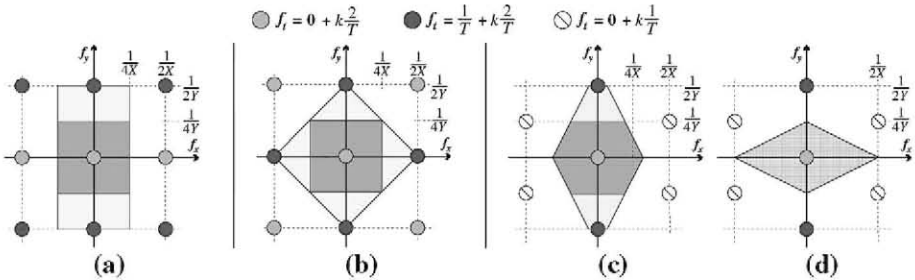


Fig. 10 Purely spatial frequencies of the: **a** horizontal Λ_{iH} , **b** BCO quincunx Λ_{iB} , and **c–d** LQ quincunx Λ_{iL} interlaced format subsampling schemes. In **a**, **b** and **c**, the *shadowed* region shows the cut of the 3D Voronoi region at $f_i=0$, whereas the *dark-grey area* depicts the Voronoi region if intra-field-only processing is imposed. In **d**, the *textured* region represents another valid unit cell, different from the Voronoi cell, for frequency preservation for the LQ quincunx scheme Λ_{iL}

intra-field processing only: as a direct result of that restriction, the preservable purely-spatial frequency ranges of the three structures diminish, as depicted in Fig. 10.

Additionally, taking as optimal the Voronoi cell of the intra-field restricted structures, the preservable frequency range of the horizontal and the BCO lattices coincide exactly (see Fig. 10a, b). As we will base the design of their corresponding filters on that specific cell, both schemes will be, in practice, completely equivalent. Therefore, we will address jointly both lattices in the design step of Sect. 4.

As for the LQ scheme, however, we will not base the design of its filter on the Voronoi cell obtained as a result of the intra-field restriction (dark-grey region of Fig. 10c). We will use instead a different unit cell, depicted in Fig. 10d, which is also perfectly valid for the reciprocal lattice of Λ_{iL} . This choice is based on two main aspects:

- According to different perceptual tests, purely horizontal and vertical frequencies have greater visual importance than diagonal variations (Sullivan et al. 1991). The selected unit cell can ideally preserve the same vertical frequency range as the Voronoi cell, but a greater horizontal span (see Fig. 10c, d).
- The specific shape of that cell eases the design of its corresponding filters, as it is possible to identify it with a linear transform of a more convenient region (see Sect. 4).

4 Filter design proposal

For complexity and compatibility reasons, we will design filters considering only information corresponding the current time step: that is, intra-frame information for progressive formats, and intra-field for interlaced schemes. Thus, our filter design will be restricted to 2D, corresponding to the spatial dimensions, and any reference to time dimensions will be completely omitted.

Let us assume a certain original lattice Λ and a certain subsampling structure Λ_S such that $\Lambda_S \subset \Lambda$. As explained in Dubois (1985) and Oppenheim and Schaffer (2009), ideal low-pass filters for both anti-aliasing and interpolation defined on Λ can be obtained by setting their spectral response to 1 inside a certain unit cell of the reciprocal subsampling structure Λ_S^* and 0 outside. That unit cell can be the Voronoi region of the reciprocal scheme, which would be related to an isotropic consideration of all the possible spatial directions Dubois (1985), but not necessarily. However, filters so constructed are not realizable in practice, as they correspond to Infinite Impulse Response (IIR) filter spatial masks and only Finite Impulse Response (FIR) spatial filters can be implemented. Although ideal spatial filters are only an unreachable mathematical entity, they do represent a useful tool for FIR filter design.

Stemming from the previous considerations, our filter design consists of the following steps: (1) a certain unit cell of the reciprocal subsampling structure Λ_S^* is chosen according to certain spectral criteria; (2) the ideal IIR filter $h(\mathbf{p})$, where $\mathbf{p} = [x, y]^T$, corresponding with the unit cell is calculated; (3) the resulting IIR filter is transformed into an FIR version $\tilde{h}(\mathbf{p})$ by using a window mask with the desired size: specifically, we will use Hamming windows, designed to maximize the side-lobe level (Theußl et al. 2000); and (4) two final corrections to that windowed version $\tilde{h}(\mathbf{p})$ are applied, obtaining $\hat{h}(\mathbf{p})$ as a final result. These proposed corrections have been proposed taking into account the specific characteristics of the digital video signal, and they allow a satisfying quality of the reconstructed sequences which would be otherwise impossible for small filter masks. These corrections are two:

- For correct video perception, it is essential to maintain exactly the mean value of the signal. To ensure that behavior, the response $\hat{H}(\mathbf{f})$ of the final filter at frequency $\bar{\mathbf{0}}$ shall be forced exactly to 1, which translates directly into

$$\hat{H}(\bar{\mathbf{0}}) = \sum_{\mathbf{p} \in \Lambda} \hat{h}(\mathbf{p}) e^{-j2\pi\bar{\mathbf{0}}^T \mathbf{p}} = \sum_{\mathbf{p} \in \Lambda} \hat{h}(\mathbf{p}) = 1 \quad (4)$$

- The continuous component of the signal to filter will have considerable energy, due to the strictly positive nature of the video signal. Therefore, the high-frequency aliases of the original spectrum generated during the subsampling process will also present spurious high-energy components at their central frequency (Dubois 1985). Thus, to correctly eliminate the aliases, it is convenient to normalize the resulting filter $\hat{H}(\mathbf{f})$ so as to force its spectral response to be identically 0 at those specific frequencies. The omission of this normalization step results, for small mask sizes, in awkward spatial patterns in the reconstructed sequence, corresponding to the high-frequency aliases not satisfactorily attenuated. This normalization step, unlike the previous one, is not common for all the filters and must be addressed specifically for each considered structure.

4.1 Filters for the proposed progressive schemes

Progressive format videos are defined over the orthorhombic lattice Λ_O of density 1, which would be also the domain where both anti-aliasing and interpolating filters for a certain subsampling structure Λ_S are defined. However, including our preference for intra-frame-only

processing, filters will be identically 0 outside the plane $t = 0$ and thus will be in practice defined only on a 2D rectangular lattice.

As discussed previously, the first step of our filter design consists in choosing a certain unit cell \mathcal{P}_S of the reciprocal subsampling structure Λ_S^* to be the spectral support of the ideal filter to approximate. The design of the filters for the subsampling lattices Λ_{pH} , Λ_{pQ} and Λ_{pN} will be based thus on the corresponding unit cells studied and selected in Sect. 2 due to their desirable frequency-preservation properties under the imposed intra-frame-only requirement. Additionally, as the selected unit cells for the two quincunx sublattices Λ_{pQ} and Λ_{pN} are coincident, their design will be addressed jointly.

4.1.1 Design for the horizontal subsampling lattice Λ_{pH}

The chosen support \mathcal{P}_{pH} of the frequency response $H_{pH}(\mathbf{f})$, $\mathbf{f} = [f_x, f_y]^T$, of the ideal IIR filter for the horizontal lattice Λ_{pH} will be the corresponding Voronoi region depicted in Fig. 5a. Applying the Inverse Fourier Transform (Dubois 1985), and taking also into account that the density of the normalized orthorhombic lattice is 1, the ideal impulse response $h(\mathbf{p})$ of the filter will be

$$h_{pH}(\mathbf{p}) = \int \int_{\mathcal{P}_{pH}} e^{j2\pi\mathbf{r}^T\mathbf{p}} d\mathbf{f} = \left(\int_{-\frac{1}{4}}^{\frac{1}{4}} e^{j2\pi f_x x} df_x \right) \left(\int_{-\frac{1}{2}}^{\frac{1}{2}} e^{j2\pi f_y y} df_y \right), \quad (5)$$

where \mathbf{p} defined over the basic orthorhombic lattice Λ_O . Using standard results from Fourier analysis (Oppenheim and Schaffer 2009), the previous expression can be simplified to

$$h_{pH}(\mathbf{p}) = \frac{1}{2} \operatorname{sinc}\left(\frac{x}{2}\right) \operatorname{sinc}(y) = \begin{cases} \frac{1}{2} \operatorname{sinc}\left(\frac{x}{2}\right) & \text{if } y = 0 \\ 0 & \text{if } y \neq 0 \end{cases}, \quad (6)$$

as both $x, y \in \mathbb{Z}$. Additionally, even for $y = 0$, the filter mask will only be different from 0 at $x = 0$ and $x = 2k + 1$, with $k \in \mathbb{Z}$.

Once we have obtained an FIR version $\tilde{h}_{pH}(\mathbf{p})$ with the desired size of this infinite response by windowing $h_{pH}(\mathbf{p})$ (Theußl et al. 2000), we apply the corresponding posterior corrections to obtain the final filter $\hat{h}_{pH}(\mathbf{p})$ with desired spectral properties. To ensure that the mean value of the video signal is kept, (4) must simply hold. To force that the spectral response of the resulting filter is exactly 0 at the center of the created aliases ($\mathbf{f} = [\pm\frac{1}{2}, 0]^T$, see Fig. 5a), the final filter must satisfy

$$\hat{H}_{pH}\left(\pm\frac{1}{2}, 0\right) = \sum_{\forall x} \hat{h}_{pH}(x) e^{\pm j2\pi(\frac{x}{2})} = \sum_{\forall x} \hat{h}_{pH}(x) (e^{\pm j\pi})^x = 0. \quad (7)$$

Using that $e^{\pm j\pi} = -1$, the above condition can be written as

$$\sum_{\forall x} \hat{h}_{pH}(x) (-1)^x = 0. \quad (8)$$

Both normalizing conditions can be simultaneously fulfilled using a simple system of equations.

4.1.2 Design for the quincunx subsampling lattices Λ_{pQ} and Λ_{pN}

The chosen support \mathcal{P}_{pQ} of the frequency response $H_{pQ}(\mathbf{f})$, $\mathbf{f} = [f_x, f_y]^T$ of the ideal IIR filter for both the stationary and non-stationary quincunx sublattices Λ_{pQ} and Λ_{pN} will be the Voronoi region depicted in Fig. 5b, c. This specific unit cell was selected mainly because of its preservation of pure horizontal and vertical frequencies.

We will apply again the Inverse Fourier Transform on the spectral response just discussed to get the IIR version of the filter, as we did previously for the horizontal sublattice Λ_{pH} . In that case, the calculation is not straightforward, as the sides of the desired spectral support \mathcal{P}_{pQ} of $H_{pH}(\mathbf{f})$ are not parallel to the coordinate axes f_x and f_y : therefore, to simplify the integration, we use an auxiliary change of variables $\mathbf{g} = \mathbf{R}\mathbf{f}$, where \mathbf{R} is a matrix representing a rotation of 45° (and therefore $|\mathbf{R}| = 1$ and $\mathbf{R}^T = \mathbf{R}$). That change of variables transforms the integration region into a different one, \mathcal{P}_R , which is square, with sides parallel to the axes and length $1/\sqrt{2}$. Then, we can write

$$h_{pQ}(\mathbf{p}) = \int \int_{\mathcal{P}_R} e^{j2\pi\mathbf{g}^T(\mathbf{R}^{-T}\mathbf{p})} d\mathbf{g} = \left[\int_{-\frac{1}{2\sqrt{2}}}^{\frac{1}{2\sqrt{2}}} \int_{-\frac{1}{2\sqrt{2}}}^{\frac{1}{2\sqrt{2}}} e^{j2\pi\mathbf{g}^T\mathbf{r}} d\mathbf{g} \right]_{\mathbf{r}=\mathbf{R}^{-T}\mathbf{p}}. \quad (9)$$

Using known Fourier analysis identities, this expression can be more clearly rewritten as

$$h_{pQ}(\mathbf{p}) = \left[\frac{1}{2} \operatorname{sinc}\left(\frac{r_1}{\sqrt{2}}\right) \operatorname{sinc}\left(\frac{r_2}{\sqrt{2}}\right) \right]_{\mathbf{r}=\mathbf{R}^{-T}\mathbf{p}} = \frac{1}{2} \operatorname{sinc}\left(\frac{x-y}{2}\right) \operatorname{sinc}\left(\frac{x+y}{2}\right), \quad (10)$$

where $x, y \in \mathbb{Z}$. Due to the specific properties of the $\operatorname{sinc}(x)$ function, the resulting response will be identically 0 for $(x \pm y) = 2k$, with $k \in \mathbb{Z}$. That means than almost half the samples can be completely ignored, even after the windowing process for obtaining the corresponding FIR version $\tilde{h}_{pQ}(\mathbf{p})$.

Again, the two discussed normalization conditions must be applied to obtain the final filter $\hat{h}_{pQ}(\mathbf{p})$. On the one hand, the result (4) must be satisfied to keep invariant the mean value of the video signal. On the other hand, the spectral response of the filter must be identically 0 at the central frequency of the aliases of the quincunx lattices, which are located at $\mathbf{f} = [\pm\frac{1}{2}, \pm\frac{1}{2}]^T$ (see Fig. 5b, c). Then, using again that $e^{\pm j\pi} = -1$, we can write

$$\hat{H}_{pQ}\left(\pm\frac{1}{2}, \pm\frac{1}{2}\right) = \sum_{\forall x,y} \hat{h}_{pQ}(x) (-1)^{(x+y)} = 0. \quad (11)$$

4.2 Filters for the proposed interlaced schemes

Interlaced format videos are defined over a vertically aligned sampling structure Λ_I of density 1, which would be also the domain where the filters of each subsampling structure Λ_S are defined. However, imposing intra-field-only processing, filters will be identically 0 outside the plane $t = 0$ of the original lattice Λ_I and therefore will be only on a 2D rectangular lattice.

The unit cells chosen for reciprocal of each of the considered subsampling structures Λ_{iH} , Λ_{iB} and Λ_{iL} are those shown in Fig. 10a, b, d. The first two unit cells, corresponding to the horizontal Λ_{iH} and BCO Λ_{iB} subsampling lattices, are coincident and are actually Voronoi cells (considering only spatial dimensions); therefore, the filters for both structures will be identical and thus addressed jointly. Whereas, the selected unit cell for the subsampling

structure LQ Λ_{iL} is not the Voronoi cell: this unit cell have been chosen to maximize the range of purely horizontal and vertical frequencies kept.

4.2.1 Design for the horizontal Λ_{iH} and BCO Λ_{iB} subsampling lattices

The chosen support \mathcal{P}_{iH} of the frequency response $H_{iH}(\mathbf{f})$, $\mathbf{f} = [f_x, f_y]^T$ of the ideal IIR filter for both the horizontal Λ_{iH} and BCO Λ_{iB} subsampling lattices will be the Voronoi region marked in dark-grey in Fig. 10a, b.

Analogously to how we proceeded for the progressive video sublattices, we apply the Inverse Fourier Transform and simplify the resulting expressions, obtaining

$$h_{iH}(x, y) = \begin{cases} \frac{1}{2} \operatorname{sinc}\left(\frac{x}{2}\right) & \text{if } y = 0 \\ 0 & \text{if } y \neq 0 \end{cases}, \quad (12)$$

where $x \in \mathbb{Z}$. That expression coincides exactly with the filter proposed for the horizontal subsampling lattice for progressive video.

The windowing of the IIR response is identical to that of the progressive format, and yields the FIR version $\hat{h}_{iH}(\mathbf{p})$ of the filter. In order to obtain the normalized final version $\hat{h}_{iH}(\mathbf{p})$ we have to apply (4) and correctly eliminate the central frequency of the generated aliases, which are located in $\mathbf{f} = [\pm\frac{1}{2}, 0]^T$ (see Fig. 10a, b). Using again the identity $e^{\pm j\pi} = -1$, the latter condition can be written as

$$\sum_{\forall x} \hat{h}_{iH}(x) (-1)^x = 0, \quad (13)$$

which coincides exactly with the expression (15) for the progressive horizontal sublattice Λ_{pH} .

4.2.2 Design for the line-quincunx subsampling scheme Λ_{iL}

The selected support \mathcal{P}_{iL} of the frequency response $H_{iH}(\mathbf{f})$ of the ideal IIR filter is the textured rhombus depicted in Fig. 10c. Analogously to the procedure used for the quincunx lattices for progressive format, we will apply a change of variables to transform that region into a simple polygon with sides aligned along the coordinate axes: now, instead of a rotation, we apply the transform $\mathbf{g} = \mathbf{A}\mathbf{f}$, where \mathbf{A} is the composition of a rotation of 45° and a scaling $\times 2$ along the f_x axis. After some calculations, and writing the positions of the interlaced sampling structure Λ_I as $\mathbf{p} = [x, y]^T = [n, 2m]^T$, with $n, m \in \mathbb{Z}$, we obtain

$$h_{iL}(n, m) = \frac{1}{2} \operatorname{sinc}\left(\frac{n-m}{2}\right) \operatorname{sinc}\left(\frac{n+m}{2}\right). \quad (14)$$

Note that this expression is similar to that obtained for the progressive quincunx lattices (10), but defined in terms of the ‘‘intra-field’’ coordinates n and m instead of x and y .

As we did in the previously addressed filters, we obtain the corresponding FIR version $\hat{h}_{iL}(\mathbf{p})$ by windowing the IIR response $h_{iL}(\mathbf{p})$. The normalized final version $\hat{h}_{iL}(\mathbf{p})$ is obtained by, first, applying (4) to forcing a correct mean value, and second, by ensuring a correct elimination of the aliases. Those aliases appear, in this case, at $\mathbf{f} = [\pm\frac{1}{2}, \pm\frac{1}{4}]^T$ (see Fig. 10d), and then the condition can be written as

$$\hat{H}_{iL}\left(\pm\frac{1}{2}, \pm\frac{1}{4}\right) = \sum_{\forall n, m} \hat{h}_{iL}(x) (-1)^{(n+m)} = 0. \quad (15)$$

5 Results and evaluation of the proposed schemes and filters

The capabilities of the considered subsampling schemes for both progressive and interlaced stereo video have been tested by comparing the objective fidelity of the subsampled/reconstructed versions with respect to the corresponding original sequence, measured in terms of *Peak Signal-to-Noise Ratio* (PSNR). Although PSNR is widely used to evaluate video quality, *Structural Similarity* (SSIM) has also been analyzed in all cases in order to confirm the observed results from a different point of view, more consistent in principle with human eye perception (Wang and Bovik 2009).

Progressive and interlaced test sequences have been processed using the filters proposed, respectively, in Sects. 4.1 and 4.2. Tests have considered multiple mask sizes in order to compare different schemes and filters in a certain range of operating conditions. To ease relevant comparisons, result display has been arranged by grouping those filters with the same number of elements: that is, each row (horizontal) filter with N^2 elements has been compared with the square (quincunx) filter of size $N \times N$. In each case, the same size have been used for the anti-aliasing and interpolation filters. The proposed filters have also been compared with the well-known bicubic interpolation algorithm to evaluate their relevance as subsampling/interpolation tools for the considered schemes.

Additionally, the overall effect of the anti-aliasing filtering has been tested. Theoretically, anti-aliasing filters are required to avoid (or at least minimize) the distorting effect of high-frequency aliases of the original baseband signal (Oppenheim and Schaffer 2009). However, they also increase linear distortion as a negative side effect. The aim of our experiments is to determine which effect (positive and negative) is dominant in common video sequences for each subsampling structure. In the tests where the anti-aliasing step has been omitted the indicated filter size refers only to the interpolation filter, as no filter is used before decimation.

The evaluation of the progressive subsampling lattices and filtering combinations have been performed on multiple sequences of different resolutions and characteristics from the Xiph.org repository (Xiph.org Test Media repository, <http://media.xiph.org/video/derf/>). For clarity, results for only three different sequences used in MPEG for 3D video coding evaluation (ISO/IEC JTC1/SC29/WG11 2010) are displayed and discussed. In Fig. 11, PSNR reconstruction quality for the three sequences under different subsampling structures and filter sizes is presented, allowing an intuitive comparison of the considered schemes. The obtained results show that, although the horizontal scheme Λ_{pH} yields better results than the quincunx structures if the anti-aliasing step is included, quincunx lattices Λ_{pQ} and Λ_{pN} without anti-aliasing filtering outperform all the other combinations. These results are consistent with the SSIM measures for the same sequences showed in Fig. 12, and are confirmed by similar results obtained for all the progressive format sequences evaluated.

These results are consistent with the spatial frequency range preserved by each lattice, previously studied in Sect. 2. The horizontal scheme Λ_{pH} allows to keep all the vertical frequency content of the original sequence, but only half the possible horizontal frequency range: thus, considerable aliasing of horizontal high-frequencies is undergone in certain sequences. Whereas, Λ_{pQ} and Λ_{pN} allow to preserve most purely vertical and horizontal frequency content of the original sequence, and thus only interpolation will be needed for most sequences. The need for the anti-aliasing filter in each progressive scheme is qualitatively show in Fig. 13, where a close-up of a frame with high-frequency content is observed. Here, we clearly see how the omission of the anti-aliasing step causes, for the horizontal subsampling scheme Λ_{pH} , jagged lines at nearly vertical orientations (see

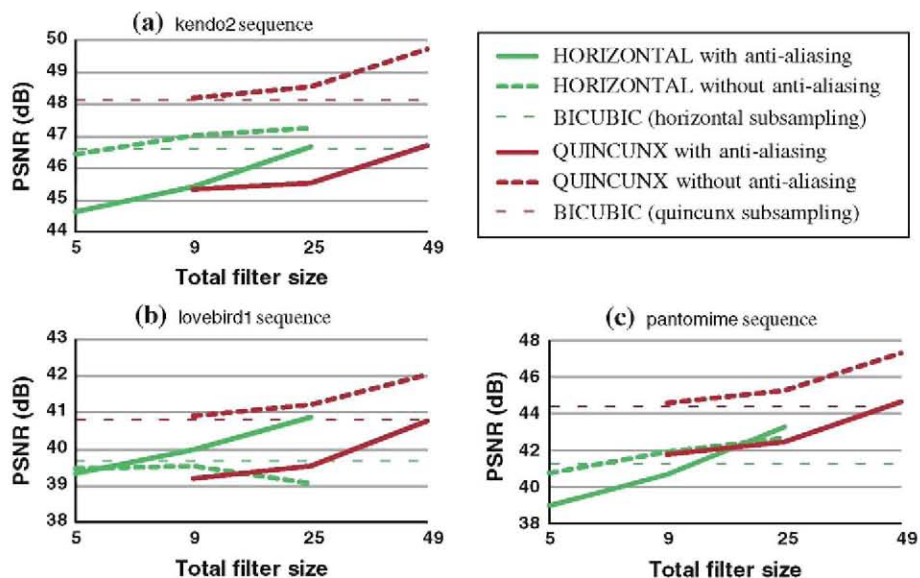


Fig. 11 Objective reconstruction quality results (PSNR) for three different progressive video sequences

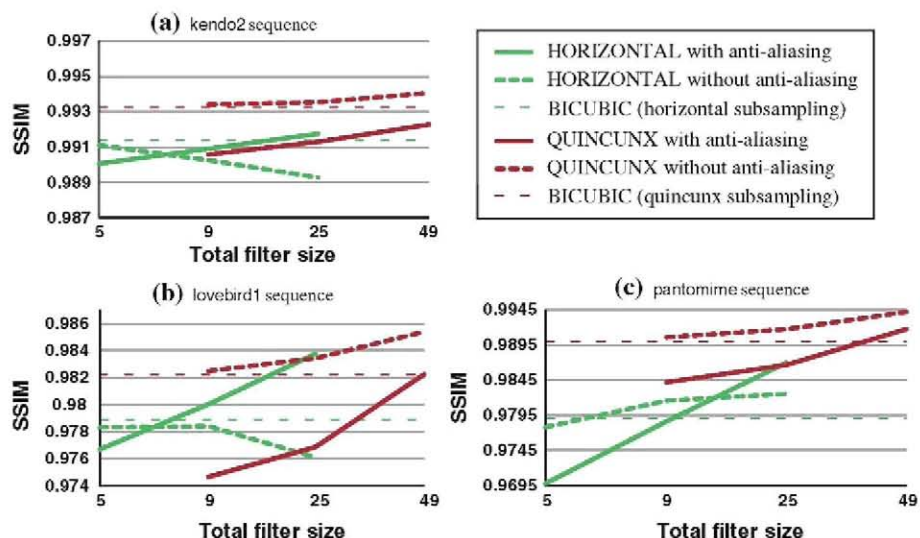


Fig. 12 Structural similarity quality results (SSIM) for three different progressive video sequences

Fig. 13c), which are avoided if the anti-aliasing filter is included (see Fig. 13d). Whereas, these incorrect patterns are not observed in the reconstructed versions corresponding to the quincunx lattices Λ_{pQ} and Λ_{pN} , either including the anti-aliasing step or, more interestingly, omitting it.

In addition, bicubic interpolation yields similar results to those provided by the smallest proposed filters for both horizontal and quincunx schemes, as showed in Figs. 11 and 12.

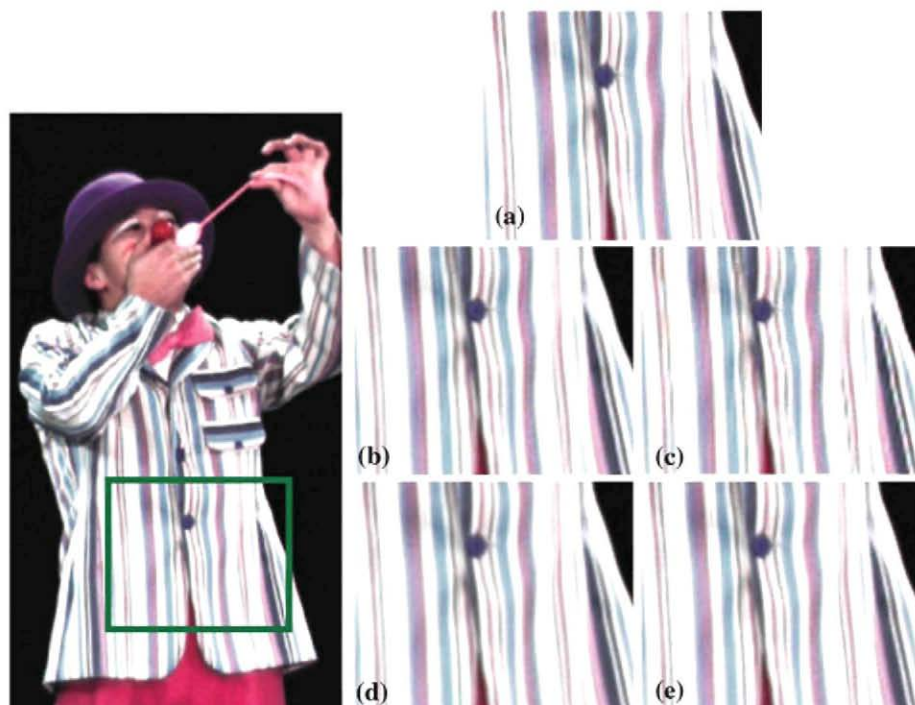


Fig. 13 Effect of the anti-aliasing step on a high-frequency detail of the pantomime progressive video sequence: **a** Original detail, reconstructed versions for the horizontal scheme Λ_{pH} **b** with and **c** without anti-aliasing filtering, and reconstructed versions for the quincunx lattices Λ_{pQ} and Λ_{pN} **d** with and **e** without anti-aliasing filtering. All reconstructed versions have been generated with their corresponding filter with size 25

However, the proposed procedure for filter mask generation proves to be more flexible and powerful, as it allows the creation of filters with the desired size, outperforming clearly the bicubic interpolation for mask sizes >9 .

As for the interlaced video format, the evaluation of the considered subsampling schemes and the filtering combinations have been performed on multiple sequences available in the Xiph.org repository (Xiph.org Test Media repository, <http://media.xiph.org/video/derf/>), and also on sequences proposed by ITU-R for perceptual evaluation (ITU-R Broadcasting Service 1994). Figure 14 presents the PSNR reconstruction quality results for three different sequences from the latter database, which shows that the horizontal Λ_{iH} and BCO Λ_{iB} lattices outperform the LQ scheme Λ_{iL} for the proposed filters and size ranges tested. Additionally, anti-aliasing filtering seems necessary for the Λ_{iH} and Λ_{iB} schemes, as the horizontal frequency content is drastically reduced. Again, these conclusions are consistent with the SSIM results obtained for the same sequences (shown in Fig. 15), and also with the results observed for the rest of evaluated sequences.

As in the progressive format case, bicubic interpolation yields for interlaced video similar results to those provided by the smallest proposed filters for both horizontal and LQ schemes, as showed in Figs. 14 and 15. But, again, as the proposed framework is valid for different filter sizes, it allows to outperform bicubic interpolation for masks with a certain number of elements.

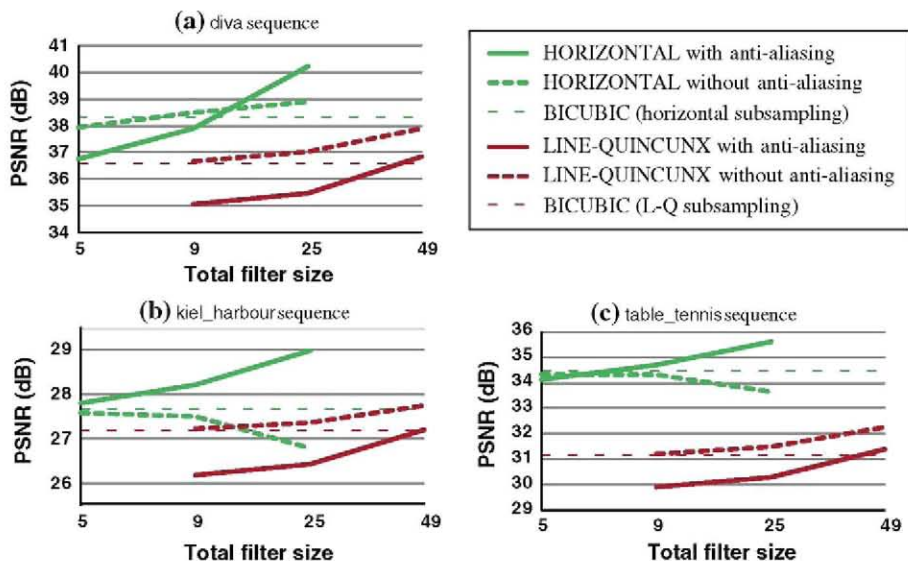


Fig. 14 Objective reconstruction quality results (PSNR) for three different interlaced video sequences

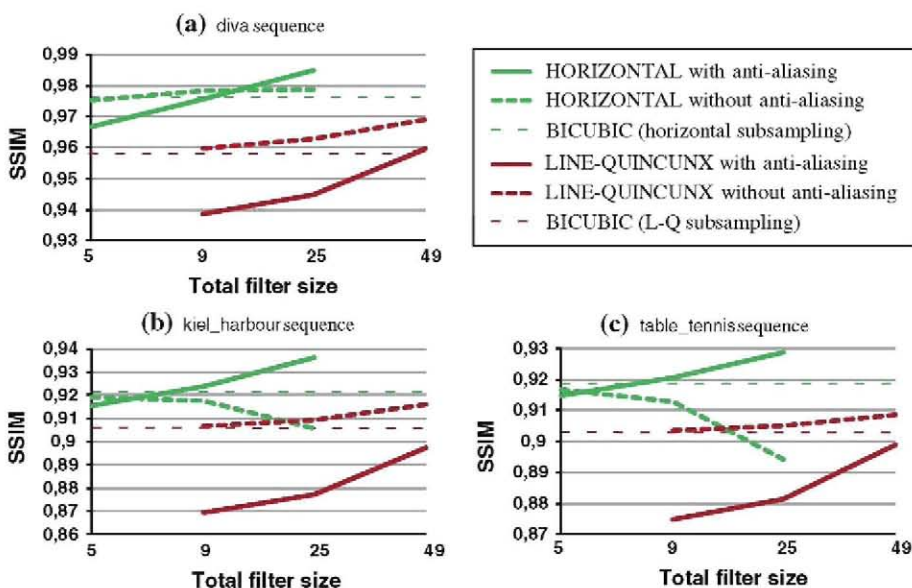


Fig. 15 Structural similarity quality results (SSIM) for three different interlaced video sequences

6 Conclusions

In this paper, a variety of frame-compatible schemes for both progressive and interlaced stereo 3DTV have been presented and studied. These schemes, specially well suited for their integration with current 2DTV equipment, are based in linear subsampling and reconstruction

techniques: therefore, their theoretical capabilities in terms of frequency preservation have been analyzed. Additionally, specific anti-aliasing and interpolation filters for the studied schemes have been specifically proposed. In their design, intra-frame-only filters for progressive and intra-field for interlaced formats has been imposed, so as to simplify as much as possible the associated processing.

Multiple tests for the proposed progressive and interlaced schemes and filters have been conducted. Results corresponding to progressive video format suggest that, under our intra-frame-only processing assumption, quincunx schemes are preferable to horizontal subsampling lattices: this is partly because the former allow to omit the anti-aliasing step without suffering incorrect high-frequency replication. Whereas, for interlaced video format with intra-field-only processing, horizontal sublattices do seem to outperform quincunx schemes.

Acknowledgments This work has been partially supported by SAPEC and the Spanish Administration agency CDTI under project CENT-VISION 2007-1007, by the Ministerio de Ciencia e Innovación of the Spanish Government under project TEC2010-20412 (Enhanced 3DTV), and by the Ministerio de Industria under project ITEA2-JEDI. Also, R. Mohedano wishes to thank the Comunidad de Madrid for a personal research grant.

References

- Díez-Ajenjo, M. A., & Capilla, P. (2010). Spatio-temporal contrast sensitivity in the cardinal directions of the colour space. A review. *Journal of Optometry*, 3(1), 2–19.
- Dubois, E. (1985). The sampling and reconstruction of time-varying imagery with application in video systems. *Proceedings of the IEEE*, 73(4), 502–522.
- Fairchild, M. D. (2005). *Color appearance models* (2nd ed.pp. 26–30). Chichester, UK: Wiley-IS&T.
- HDMI Licensing, LLC. (May 2009). High definition multimedia interface: Specification version 1.4a.
- ISO/IEC JTC1/SC29/WG11. (2010). Draft call for proposals on 3D video coding technology, Output Doc. N11679.
- ITU-R Broadcasting Service. (1994). Test pictures and sequences for subjective assessments of digital codecs conveying signals produced according to Rec. ITU-R BT.601, Recommendation ITU-R BT.802-1.
- ITU-R Broadcasting Service. (2002). Methodology for the subjective assessment of quality of television pictures, Recommendation ITU-R BT.500-11.
- ITU-R Broadcasting Service. (2009). Features of three-dimensional television video systems for broadcasting, report ITU-R BT.2160.
- ITU-T and ISO/IEC JTC1. (2010). Advanced video coding for generic audiovisual services, ITU-T recommendation H.264 and ISO/IEC 14496-10 (MPEG-4 AVC).
- Kelly, D. H. (1979). Motion and vision II: Stabilized spatio-temporal threshold surface. *Journal of the Optical Society of America*, 69(10), 1340–1349.
- Kelly, D. H. (1983). Spatiotemporal variation of chromatic and achromatic contrast thresholds. *Journal of the Optical Society of America*, 73(6), 742–749.
- King, P. (2011). *Global 3D-enabled device sales forecast*. Strategy Analytics Inc. <http://www.strategyanalytics.com/default.aspx?mod=reportabstractviewer&a0=6153>
- Mannos, J. L., & Sakrison, D. J. (1974). The effects of a visual fidelity criterion on the encoding of images. *IEEE Transactions on Information Theory*, 20(4), 525–536.
- Merkle, P., Muller, K., & Wiegand, T. (2010). 3D Video: Acquisition, coding, and display. *IEEE Transactions on Consumer Electronics*, 56(2), 946–950.
- Nintendo®. (2011). *Nintendo 3DS™ hardware specifications*. March 27, 2011. <http://www.nintendo.com/3ds/hardware/specs>
- Oppenheim, A. V., & Schaffer, R. W. (2009). *Discrete-time signal processing*. New Jersey, USA: Prentice Hall.
- Sullivan, J. R., Ray, L. A., & Miller, R. (1991). Design of minimum visual modulation halftone patterns. *IEEE Transactions on Systems, Man and Cybernetics*, 21(1), 33–38.
- Theußl, T., Hauser, H., Gröller, E., Sullivan, J. R., Ray, L. A., & Miller, R. (2000). Mastering windows: Improving reconstruction. *IEEE/ACM SIGGRAPH Symposium on Volume Visualization*, pp. 101–108.

- Vetro, A., Tourapis, A. M., Muller, K., & Chen, T. (2011). 3D-TV content storage and transmission. *IEEE Transactions on Broadcasting*, 57(2(Part 2)), 384–394.
- Wang, Z., & Bovik, A. C. (2009). Mean squared error: Love it or leave it?. *IEEE Signal Processing Magazine*, 26(1), 98–117.

Author Biographies



Raúl Mohedano received the Ingeniero de Telecomunicación degree (5 years engineering program) in 2006 from the Universidad Politécnica de Madrid (UPM), Madrid, Spain. Since 2007 he has been a member of the Grupo de Tratamiento de Imágenes (Image Processing Group) of the UPM, where he holds a personal research grant from the Comunidad de Madrid. His research interests are in the area of computer vision.



Carlos Muñoz received the Ingeniero de Telecomunicación degree (6 years engineering program) in 1982 and the Doctor Ingeniero de Telecomunicación degree (PhD in Communications) in 1986, both from the Universidad Politécnica de Madrid (UPM), Madrid, Spain. Since 1982 he is a member of the faculty of the UPM, where is currently Associate Professor of Signal Theory and Communications. His professional and research interests are in the areas of digital communications and image and video compression. He has been actively involved in European projects (Eureka, Acts, IST) and national projects.



Julián Cabrera received the Ingeniero de Telecomunicación degree (6 years engineering program) in 1996 and the Doctor Ingeniero de Telecomunicación degree (PhD in Communications) in 2003, both from the Universidad Politécnica de Madrid (UPM), Madrid, Spain. Since 1996 he is a member of the Image Processing Group of the UPM. He was a PhD scholar of the Information Technology and Telecommunication Programs of the Spanish National Research Plan from 1996 till 2001. Since 2001 he is a member of the faculty of the UPM, and since 2003 he is an Associate Professor of Signal Theory and Communications. His professional interests include image and video coding, design and development of multimedia communications systems, focusing on Multiview Video Coding (MVC), 3D Video Coding and video transmission over variable rate channels. He has been actively involved in European projects (Acts, Telematics, IST) and national projects.



Fernando Jaureguizar received the Ingeniero de Telecomunicación degree (6 years engineering program) in 1987 and the Doctor Ingeniero de Telecomunicación degree (PhD in Communications) in 1994 (Doctoral Graduation Award), both from the Universidad Politécnica de Madrid (UPM), Madrid, Spain. Since 1987 he is a member of the Image Processing Group of the UPM. He was a PhD scholar of the Information Technology and Telecommunication Programs of the Spanish National Research Plan from 1988 till 1991. Since 1991 he is a member of the faculty of the UPM, and since 1995 he is an Associate Professor of Signal Theory and Communications. His professional interests include digital image processing, video coding, 3DTV, and design and development of multimedia communications systems. He has been actively involved in European projects (Eureka, Acts, IST) and national projects.



Narciso García received the Ingeniero de Telecomunicación degree (5 years engineering program) in 1976 (Spanish National Graduation Award) and the Doctor Ingeniero de Telecomunicación degree (PhD in Communications) in 1983 (Doctoral Graduation Award), both from the Universidad Politécnica de Madrid (UPM), Madrid, Spain. Since 1977 he is a member of the faculty of the UPM, where is currently Professor of Signal Theory and Communications. He leads the Grupo de Tratamiento de Imágenes (Image Processing Group) of the UPM. He was Area Coordinator of the Spanish Evaluation Agency from 1990 to 1992 and evaluator, reviewer, and auditor of European programs since 1990. His professional and research interests are in the areas of digital image and video compression and of computer vision. He has been actively involved in European projects (Eureka, Race, Cost, Esprit, Acts, Telematics, IST) and was a co-writer of EBU proposal, base of the ITU standard for digital transmission of TV at 34–45 Mb/s (ITU-T J.81).