

# ANALYSIS OF THE ALTERNATIVES FOR STEREO MULTIPLEX IN 3DTV BROADCAST DELIVERY

Raúl Mohedano, Julián Cabrera, Fernando Jaureguizar, Narciso García

Grupo de Tratamiento de Imágenes, Universidad Politécnica de Madrid, 28040 Madrid, Spain  
 {rmp, julian.cabrera, fjn, narciso}@gti.ssr.upm.es; www.gti.ssr.upm.es

## ABSTRACT

Although 3DTV has led the evolution of television market, its delivery by broadcast networks is still small. Now, 3DTV transmissions are usually done by combining both views into one common frame (side by side) to be able to use standard HDTV transmission equipment. Today, orthogonal subsampling is mostly used, but other alternatives will appear soon.

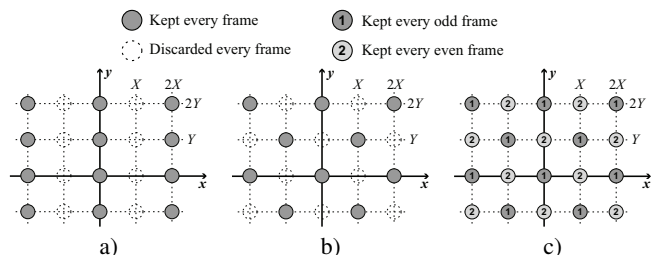
Here, different subsampling schemes for both progressive and interlaced 3DTV are considered. For each possible scheme, its preserved frequency content is analyzed and a simple interpolation filter is designed. The analysis is carried out for progressive and interlaced video and the designed filters are applied on different sequences, showing the advantages and disadvantages of the different options.

**Index Terms**— 3DTV, side-by-side, subsampling, interpolation

## 1. INTRODUCTION

Nowadays 3D video technology leads the evolution of movie theaters and has clearly set foot in the TV set consumer market. However, its delivery by broadcasting networks is still lagging behind, mainly because of the lack of completely standardized end-to-end solutions and the need for enlarged bit-streams for these signals. Therefore the usual procedure is nowadays the spatial multiplexing of both views into one to allow the use of standard HDTV transmission equipment. This procedure implies halving the samples of each view at encoding and the interpolation of those missing samples at decoding. Among different possibilities, the side-by-side multiplex of orthogonally subsampled views is the usual choice. Obviously, the huge number of TV sets endowed with HDMI 1.3 has helped that decision. Nevertheless, the adoption of HDMI 1.4 with a new set of available subsampling schemes allows different alternatives.

Here, we analyze several subsampling schemes for progressive and interlaced 3DTV and look for simple interpolation schemes, aiming for the maximum reconstruction quality within the predefined computation schemes. So, the paper analyzes the spectral implications of the different subsampling structures by studying their Voronoi cells, and designs simple purely-spatial filters fulfilling the requirements for frequency preservation, which provides satisfying reconstruction results even for small filter masks thanks to a newly proposed normalization step. This analysis is conducted initially for progressive video and later extended to interlaced video, discussing the similarities and differences between them. Then, the designed filters are tested on different sequences and the set of results is interpreted. Finally, rules on the use of subsampling schemes and interpolation filters are given.



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

## 2. LATTICES FOR PROGRESSIVE 3D VIDEO FORMAT

We will consider for study three different lattices for progressive video subsampling with density  $1/2$ , which we will denote by: horizontal  $\Lambda_{pH}$  (horizontally-aligned sampling of 2:1), stationary quincunx  $\Lambda_{pQ}$ , and non-stationary quincunx  $\Lambda_{pN}$  (referred to as FCO, Face-Centered Orthorhombic lattice, in [1]).

### 2.1. Lattice definition and layout

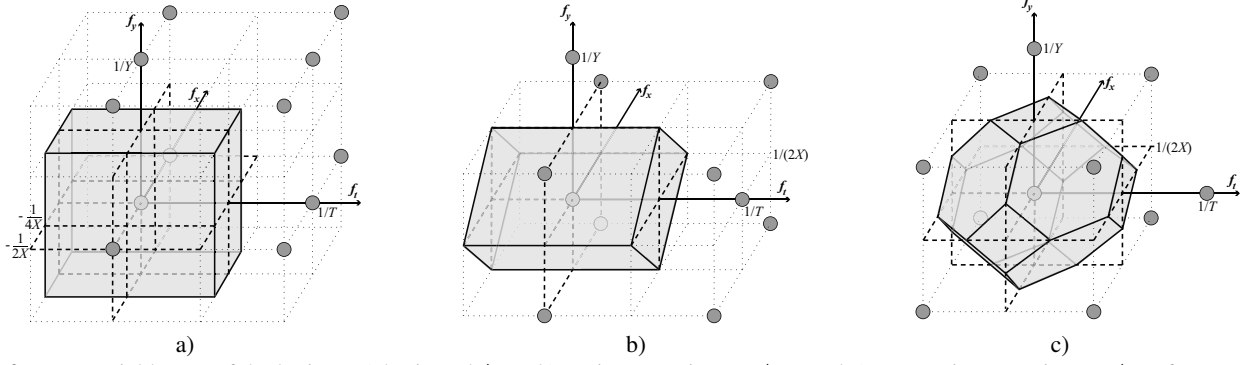
Following the notation presented by Dubois in his seminal paper [1], we will represent each lattice  $\Lambda$  by means of a matrix  $V$  whose columns contain the coordinates of the vectors of its generating basis  $B = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$  with respect to the canonical basis for  $\mathbb{R}^3$ . Here, the  $x$  and  $y$  components represent, respectively, the horizontal and vertical directions, and  $t$  the temporal evolution. The corresponding reciprocal lattice could be analogously represented by a matrix  $U$  such that  $U^T V = I$ , where  $I$  represents the identity matrix.

The horizontal  $\Lambda_{pH}$ , stationary quincunx  $\Lambda_{pQ}$ , and non-stationary quincunx  $\Lambda_{pN}$  subsampling lattices are described by the matrices

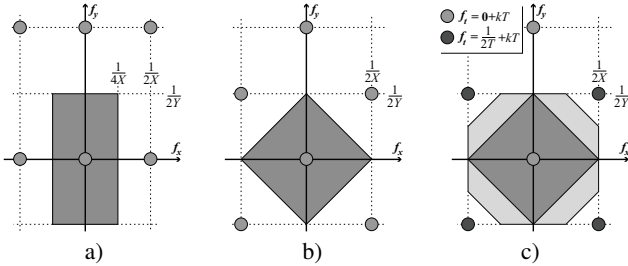
$$V_{pH} = \begin{pmatrix} 2X & 0 & 0 \\ 0 & Y & 0 \\ 0 & 0 & T \end{pmatrix}, V_{pQ} = \begin{pmatrix} 2X & X & 0 \\ 0 & Y & 0 \\ 0 & 0 & T \end{pmatrix}, V_{pN} = \begin{pmatrix} 2X & X & X \\ 0 & Y & 0 \\ 0 & 0 & T \end{pmatrix}, \quad (1)$$

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. In general, we could consider square pixels and thus  $X = Y$ . The spatial-temporal layout of these subsampling lattices is shown in Fig. 1, and their corresponding reciprocal lattices and Voronoi regions are depicted in Fig. 2.

Although the three subsampling lattices have the same overall density  $1/2$ , they present different frequency characteristics. First, the efficiency of a lattice is related to its ability to pack non-overlapping spheres centered at the points of reciprocal lattice [1], so the non-stationary quincunx lattice seems to show the best characteristics of all in terms of the potential capability of frequency preser-



**Fig. 2.** Frequential layout of the lattices: a) horizontal  $\Lambda_{pH}$ , b) stationary quincunx  $\Lambda_{pQ}$  and c) non-stationary quincunx  $\Lambda_{pN}$  for progressive video subsampling. In each case, the Voronoi region has been depicted as a translucent grey polyhedron.



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

vation, followed by the stationary quincunx. However, the human visual system is much more sensitive to horizontal and vertical spatial frequencies than to diagonal variations [2]. Then, the perceptual difference between the non-stationary and the stationary quincunx lattices can be lower than geometrically expected (see Fig. 3).

In addition, unlike  $\Lambda_{pH}$  and  $\Lambda_{pQ}$ , the non-stationary quincunx lattice  $\Lambda_{pN}$  do not remain invariant for every frame. This is reflected in its Voronoi region (see Fig. 2.c), which cannot be generated by translating a polygon along the  $f_t$  axis. It means that it is not possible to construct a filter for ideal subsampling and reconstruction of the original resolution preserving the frequencies within that region without considering information of multiple frames. If we impose intra-frame-only processing to simplify the resulting down/upsampling system, the desirable preservable frequency region for the non-stationary quincunx lattice will coincide with the Voronoi region of the stationary quincunx (see Fig. 3.b) and c). Thus we will address jointly both quincunx lattices  $\Lambda_{pH}$  and  $\Lambda_{pQ}$ , and we will not explicitly include the time-related dimensions  $t$  and  $f_t$ .

## 2.2. Filter design proposal

As we have stated before, we will construct filters using only intra-frame information: thus, we will restrict our design to 2D, that is, the spatial dimensions. We will assume, for simplicity, square pixels and a normalized orthorhombic lattice  $\Lambda_O$ : that is,  $X = Y = 1$ .

The ideal low-pass filters for anti-aliasing and interpolation coincide [3], and their spectral support is the Voronoi region of the subsampling lattice. Therefore, the proposed filters have been designed following three steps. First, the ideal (Infinite Impulse Response) IIR filter  $h(\mathbf{p})$ , where  $\mathbf{p} = [x, y]^T$ , corresponding with the Voronoi region is calculated. Second, the result is transformed into a FIR (Finite Impulse Response) filter  $\tilde{h}(\mathbf{p})$  by windowing the IIR version

to the desired size: specifically, we will use Hamming windows, designed to maximize the side-lobe level [4]. And finally two final corrections to that windowed versions  $\tilde{h}(\mathbf{p})$  have been proposed, allowing a satisfying quality of the reconstructed sequences which would be otherwise impossible for small filter masks:

1. To maintain exactly the mean tone of the video sequence, the response  $\tilde{H}(\bar{\mathbf{0}})$  of the resulting filter at frequency  $\bar{\mathbf{0}}$  will be forced exactly to 1. This condition translates directly into

$$\tilde{H}(\bar{\mathbf{0}}) = \sum_{\mathbf{p} \in \Lambda_O} \tilde{h}(\mathbf{p}) e^{-j2\pi\bar{\mathbf{0}}^T \mathbf{p}} = \sum_{\mathbf{p} \in \Lambda_O} \tilde{h}(\mathbf{p}) = 1 \quad (2)$$

2. Due to the strictly positive nature of the video signal, its continuous component will be considerably powerful, and so will be the central component of the high-frequency aliases generated during the subsampling [1]. Thus, it is convenient to normalize the filters to force the spectral response of the interpolation filter to be identically 0 at those frequencies. The omission of that normalization step results, for small mask sizes, in awkward spatial patterns in the reconstructed sequence.

### 2.2.1. Horizontal subsampling lattice

The support of the frequency response  $H_{pH}(\mathbf{f})$ ,  $\mathbf{f} = [f_x, f_y]^T$ , of the ideal IIR filter for the horizontal lattice  $\Lambda_{pH}$  will be the corresponding Voronoi region depicted in Fig. 3.a), which will be denoted by  $\mathcal{P}_{pH}$ . Thus, taking into account that the density of the normalized orthorhombic lattice is 1, the ideal impulse response  $h(\mathbf{p})$  of the filter, defined over the basic orthorhombic lattice  $\Lambda_O$  will be

$$h_{pH}(\mathbf{p}) = \iint_{\mathcal{P}_{pH}} e^{j2\pi\mathbf{f}^T \mathbf{p}} d\mathbf{f} = \left( \int_{-\frac{1}{4}}^{\frac{1}{4}} e^{j2\pi f_x x} df_x \right) \left( \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{j2\pi f_y y} df_y \right). \quad (3)$$

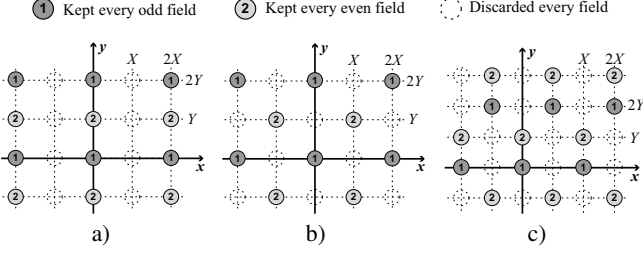
Using standard results from Fourier analysis [3], we can write

$$h_{pH}(\mathbf{p}) = \frac{1}{2} \text{sinc}\left(\frac{x}{2}\right) \text{sinc}(y) = \frac{1}{2} \text{sinc}\left(\frac{x}{2}\right) \quad (4)$$

for  $y = 0$ , and 0 for  $y \neq 0$ . 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  $h_{pH}(\mathbf{p})$  [4], we ensure 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. 3.a)). Then,

$$H_{pH}\left(\pm\frac{1}{2}, 0\right) = \sum_{\forall x} \tilde{h}_{pH}(x) e^{\pm j2\pi(\frac{x}{2})} = \sum_{\forall x} \tilde{h}_{pH}(x) (e^{\pm j\pi})^x. \quad (5)$$



**Fig. 4.** Spatial-temporal layout of the structures: a) horizontal  $\Lambda_{iH}$ , b) BCO quincunx  $\Lambda_{iB}$  and c) LQ quincunx  $\Lambda_{iL}$  for interlaced video subsampling.

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

$$\sum_{\forall x} \tilde{h}_{pH}(x) (-1)^x = 0. \quad (6)$$

### 2.2.2. Quincunx subsampling lattices

As explained before, both stationary and non-stationary quincunx schemes can preserve the same frequency ranges under the stated conditions (intra-frame processing), so we will address them jointly.

We intend to construct a filter defined on the orthorhombic lattice  $\Lambda_O$  whose frequency response  $H_{pQ}(\mathbf{f})$  is 1 within the corresponding Voronoi region, depicted in Fig. 3.b) and denoted from now on by  $\mathcal{P}_{pQ}$ , and 0 outside. As the sides of  $\mathcal{P}_{pQ}$  are not oriented as the coordinate axes  $f_x$  and  $f_y$ , we use a change of variables  $\mathbf{g} = \mathbf{R}\mathbf{f}$  to simplify the integration, where  $\mathbf{R}$  is a matrix representing a rotation of  $45^\circ$ , and therefore  $|\mathbf{R}| = 1$  and  $\mathbf{R}^{-T} = \mathbf{R}$ . That changes the integration region into a different one,  $\mathcal{P}_S$ , which is square (oriented with the axes) with sides of length  $1/\sqrt{2}$ . Now

$$h_{pQ}(\mathbf{p}) = \iint_{\mathcal{P}_S} e^{j2\pi\mathbf{g}^T(\mathbf{R}^{-T}\mathbf{p})} d\mathbf{g} = \left[ \int_{-\frac{1}{2\sqrt{2}}}^{\frac{1}{2\sqrt{2}}} \int_{-\frac{1}{2\sqrt{2}}}^{\frac{1}{2\sqrt{2}}} e^{j2\pi\mathbf{g}^T\mathbf{r}} d\mathbf{g} \right]_{\mathbf{r}=\mathbf{R}^{-T}\mathbf{p}} \quad (7)$$

and, using known Fourier identities [3], we can write

$$\begin{aligned} h_{pQ}(\mathbf{p}) &= \left[ \frac{1}{2} \operatorname{sinc}\left(\frac{r_1}{\sqrt{2}}\right) \operatorname{sinc}\left(\frac{r_2}{\sqrt{2}}\right) \right]_{\mathbf{r}=\mathbf{R}^{-T}\mathbf{p}} = \\ &= \frac{1}{2} \operatorname{sinc}\left(\frac{x-y}{\sqrt{2}}\right) \operatorname{sinc}\left(\frac{x+y}{\sqrt{2}}\right). \end{aligned} \quad (8)$$

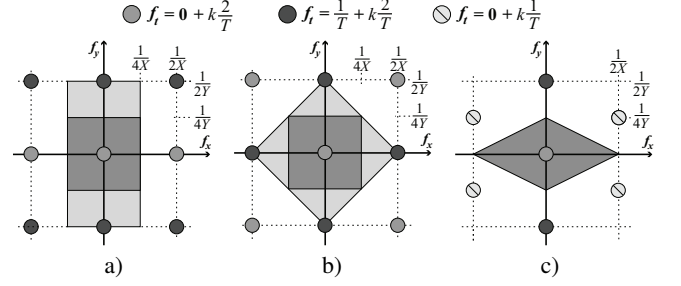
Due to the specific properties of the  $\operatorname{sinc}(x)$  function, the resulting response will be identically 0 for  $(x \pm y) = 2k$ , with  $k \in \mathbb{Z}$ . That means that almost half the samples can be completely ignored, even after the windowing process for obtaining the corresponding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

As for the post-normalization of the filter mask, the aliases of the quincunx lattices are located at  $\mathbf{f} = [\pm\frac{1}{2}, \pm\frac{1}{2}]^T$  (see Fig. 3.b) and c)). Then, using again that  $e^{\pm j\pi} = -1$ , we can write

$$H_{pQ}\left(\pm\frac{1}{2}, \pm\frac{1}{2}\right) = \sum_{\forall x,y} \tilde{h}_{pQ}(x) (-1)^{(x+y)} = 0. \quad (9)$$

## 3. LATTICES FOR INTERLACED 3D VIDEO FORMAT

We will consider three different structures for interlaced video subsampling: horizontal  $\Lambda_{iH}$ , BCO (Body-Centered Orthorhombic) quincunx  $\Lambda_{iB}$  and LQ (Line-Quincunx)  $\Lambda_{iL}$ .



**Fig. 5.** Purely spatial frequencies of the interlaced format subsampling lattices for: a) horizontal  $\Lambda_{iH}$ , b) BCO quincunx  $\Lambda_{iB}$  and c) LQ quincunx lattice  $\Lambda_{iL}$ . Shaded: cut of the 3D Voronoi region at  $f_t=0$ . Dark-grey: preserved region for intra-field-only processing.

### 3.1. Lattice definition and layout

Using the same notation as in Sec. 2, the considered structures are described by the matrices

$$\mathbf{V}_{iH} = \begin{pmatrix} 2X & 0 & 0 \\ 0 & 2Y & Y \\ 0 & 0 & \frac{T}{2} \end{pmatrix}, \mathbf{V}_{iB} = \begin{pmatrix} 2X & 0 & X \\ 0 & 2Y & Y \\ 0 & 0 & \frac{T}{2} \end{pmatrix}, \mathbf{V}_{iL} = \begin{pmatrix} 2X & X & 0 \\ 0 & 2Y & 0 \\ 0 & 0 & T \end{pmatrix}. \quad (10)$$

In fact,  $\Lambda_{iL}$  is not actually a lattice, as it is composed of two cosets [1] such that  $\mathbf{c} = [0, Y, \frac{T}{2}]^T$ . The spatial-temporal layout of the three subsampling structures is shown in Fig. 4.

A careful analysis of the Voronoi regions for the three described structures indicates that, ideally, the BCO lattice can preserve a greater range of spatial frequencies than the other two structures, keeping specially both pure horizontal and vertical frequencies (see Fig. 5). However, if we restrict the analysis to intra-field processing only, the preservable frequency ranges of the three structures diminish, and those of the horizontal and the BCO lattices coincide exactly, as depicted in Fig. 5. Thus, we will address jointly both lattices.

### 3.2. Filter design proposal

Analogously to the progressive format, the design of the anti-aliasing and interpolation filters on the interlaced lattice  $\Lambda_I$  will follow: 1) calculation of the ideal IIR filter, 2) windowing to limit its response to an FIR filter, and 3) normalization. As all the calculations are similar to those showed in Subsec. 2.2, we will only include here certain specific details of the interlaced format structures.

#### 3.2.1. Horizontal and BCO subsampling lattices

In that case, the support of the ideal IIR filter will be the Voronoi region marked in dark-grey region in Fig. 5.a) and b). After some calculations, and taking into account the lattice density, we can write

$$h_{iH}(x, y) = h_{iB}(x, y) = \frac{1}{2} \operatorname{sinc}\left(\frac{x}{2}\right) \quad (11)$$

for  $y = 0$ , and 0 for  $y \neq 0$ , which exactly coincides with the filter proposed for the horizontal subsampling lattice for progressive video. The windowing and subsequent normalization of that IIR response are identical to that of the progressive format.

#### 3.2.2. Line-quincunx subsampling structure

The support of the ideal IIR filter will be the dark-grey rhombus depicted in Fig. 5.c). As we did for the quincunx lattices for progressive format, we will apply a change of variables to transform that

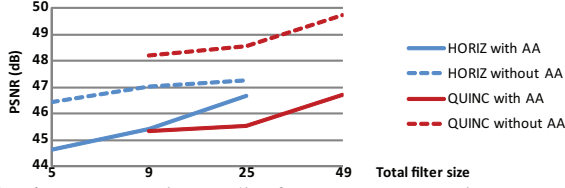


Fig. 6. Reconstruction quality for kendo2 progressive sequence.

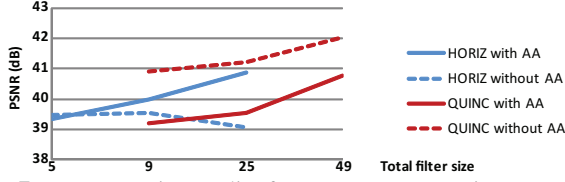


Fig. 7. Reconstruction quality for lovebird1 progressive sequence.

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 scaling  $\times 2$  along the  $f_x$  axis and a rotation of  $45^\circ$ . After some calculations, we obtain

$$h_{iL}(n, m) = \frac{1}{2} \operatorname{sinc}\left(\frac{n-m}{\sqrt{2}}\right) \operatorname{sinc}\left(\frac{n+m}{\sqrt{2}}\right), \quad (12)$$

where we have used that  $\mathbf{p} \in \Lambda_I$  can be written as  $\mathbf{p} = [x, y]^T = [n, 2m]^T$ , with  $n, m \in \mathbb{Z}$ .

Once the IIR filter has been windowed, post-normalization is performed. In that case, aliases appear at  $\mathbf{f} = [\pm\frac{1}{2}, \pm\frac{1}{4}]^T$  (see Fig. 5.c)). Proceeding as previously, we obtain the condition

$$H_{iL}\left(\pm\frac{1}{2}, \pm\frac{1}{4}\right) = \sum_{\forall n, m} \tilde{h}_{iL}(x) (-1)^{(n+m)} = 0. \quad (13)$$

#### 4. EVALUATION OF THE STUDIED ALTERNATIVES

The capabilities of the analyzed subsampling schemes have been tested by comparing the objective fidelity (PSNR) of the subsampled/reconstructed versions with respect to their original sequence. The progressive and interlaced test sequences have been processed using the filters proposed in Sec. 2 and 3. In order to compare equivalent schemes in different conditions, multiple mask sizes have been tested, grouping those filters with the same number of elements, that is, each row filter with  $N^2$  elements has been compared with the square filter of size  $N \times N$ .

In addition, although anti-aliasing (AA) filters are theoretically necessary to minimize the distorting effect of replicated frequencies [3], they also increase linear distortion as a negative side effect. Thus, the overall effect of AA filtering has been tested as well.

The schemes for progressive format have been tested on different sequences used for 3D video coding evaluation in MPEG [5]. Results (Fig. 6 and 7) indicate that: 1) quincunx lattices improve if the AA stage is omitted, whereas AA seems necessary for the horizontal scheme, and 2) although horizontal halving has better results than the quincunx structures if the AA step is included, quincunx lattices without AA filtering outperforms all the other combinations.

Regarding the schemes for interlaced format, tests have been performed on videos proposed by ITU-R for perceptual evaluation [6]. Results (Fig. 8 and 9) show that the horizontal and BCO lattices outperform the LQ scheme for the proposed filters and sizes, and anti-aliasing filtering seems necessary for the horizontal schemes, as the horizontal frequency content is drastically reduced.

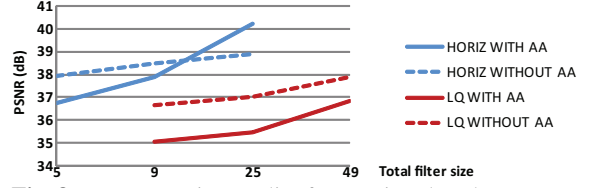


Fig. 8. Reconstruction quality for diva interlaced sequence.

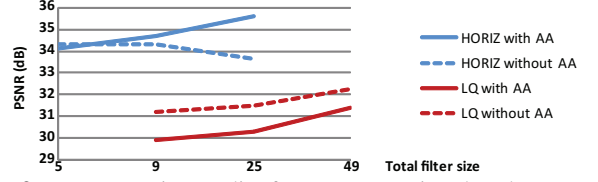


Fig. 9. Reconstruction quality for table.tennis interlaced sequence.

#### 5. CONCLUSION

The capabilities of different schemes for subsampling and reconstruction of both progressive and interlaced 3DTV have been analyzed. Their theoretical capabilities have been studied, and multiple tests for both progressive and interlaced video format have been conducted, all of them restricted to intra-frame and intra-field processing to simplify as much as possible the associated processing.

Results for progressive video format indicate that, under those test conditions, quincunx subsampling lattices (stationary or not) are preferable to horizontal resolution halving, and also suggest that quality could be improved by omitting the anti-aliasing filtering step. Whereas, horizontal halving schemes do seem to outperform other quincunx structures for interlaced formats.

However, the performed analysis of the 3D Voronoi regions of the discussed subsampling structures indicate that certain schemes (mainly those for interlaced format) could be improved by using inter-frame or inter-field processing for high-frequency preservation.

#### 6. ACKNOWLEDGMENTS

This work has been partially supported by SAPEC and the Spanish Administration agency CDTI under project CENIT-VISION 2007-1007, and by the Ministerio de Ciencia e Innovación of the Spanish Government under projects TEC2007-67764 (SmartVision) and TEC2010-20412 (Enhanced 3DTV). Also, R. Mohedano wishes to thank the Comunidad de Madrid for a personal research grant.

#### 7. REFERENCES

- [1] E. Dubois, "The sampling and reconstruction of time-varying imagery with application in video systems," *Proceedings of the IEEE*, vol. 73, no. 4, pp. 502–522, 1985.
- [2] J. R. Sullivan, L. A. Ray, and R. Miller, "Design of minimum visual modulation halftone patterns," *IEEE Trans. Systems, Man and Cybernetics*, vol. 21, no. 1, pp. 33–38, 1991.
- [3] A. V. Oppenheim and R. W. Schaffer, *Discrete-time signal processing*, Prentice Hal, New Jersey, USA, 3rd edition, 2009.
- [4] T. Theußl, H. Hauser, and E. Gröller, "Mastering windows: improving reconstruction," in *IEEE Symposium on Volume Visualization*, 2000, pp. 101–108.
- [5] ISO/IEC JTC1/SC29/WG11, "Draft Call for Proposals on 3D Video Coding Technology," Output Doc. N11679, 2010.
- [6] Rec. ITU-R BT.802-1, "Test pictures and sequences for subjective assessments of digital codecs conveying signals produced according to Rec. ITU-R BT.601," 1994.