The HAMLET Generic Format Converter

Stéphane Pigeon
UCL - Laboratoire de Télécommunications et Télédétection
Place du Levant, 2 – B–1348 Louvain-La-Neuve – Belgium
E-Mail : Pigeon@tele.ucl.ac.be

Abstract

This paper investigates the definition of a generic format converter that is accepting at its I/O both interlaced and progressive formats. These conversions make use of the general sampling theory which was proposed recently to handle interlaced images. This paper especially defines the 4:2:2 deinterlacer that has been used within the RACE2110/HAMLET project in order to study the coding efficiency of progressive sequences.

1 Introduction

This paper describes the format converter used within HAMLET to test the influence of the scanning format on the coding efficiency [1, 2]. Although the HAMLET definition of a generic format converter and the work of RACE 2055/TRANSIT, dealing with image format transcoding [3], are related, specificities of the coding application in HAMLET have been taken into account to lead to further developments. Unlike TRANSIT, the HAMLET Generic Format Converter (GFC) is basically meant to be placed in front of a coder, converting interlaced sources into progressive in order to get the best coding efficiency. To achieve this result, a particular attention must be paid to the quality of the construction of the progressive sequence from the interlaced input in order to recover the "analog" scene hidden behind the interlaced input. This can only be achieved by using a finely tuned motion estimation and compensation. The analysis inside HAMLET is based on the *general sampling theory* which was proposed recently to handle interlaced images and proved to be successful [4, 5].

As within HAMLET, we are interested in studying the coding efficiency, the GFC has to work in close relationship with a coder. Although the "generic" term suggests the format coder to work in a

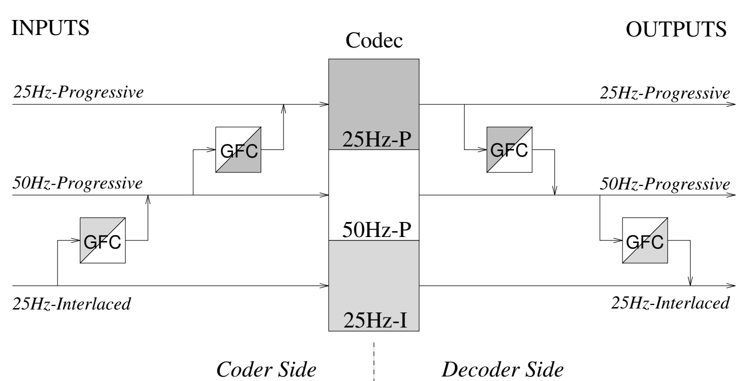

Figure 1: Study case for the generic format converter

more general context, switching directly from one format to another, a study case as represented in figure 1 has been adopted for this work since it offers the advantage to distinguish both coder and decoder sides. Indeed, as the decoder is intended to be replicated in each receiver set, economical considerations will force the market to offer, at this side, low-cost format conversions only. On the opposite, the coder side is able to deal with high-cost algorithms in order to get the best coding efficiency.

2 Generalized Sampling Theorem

From the Nyquist theorem, it is well-known that any sequence whose spectrum is limited to $1/T$ can theoretically be recovered exactly as long as it is sampled at a rate greater than $2/T$. Actually, this condition is sufficient but not necessary : the same sequence can also be recovered from N sequences at the same rate $2/NT$ but with different phases [6].

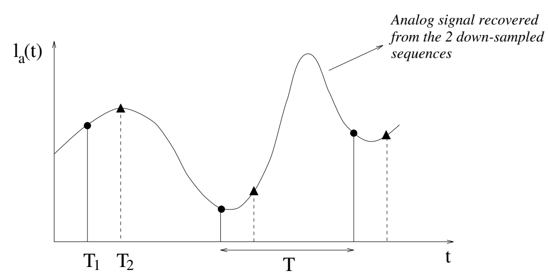

Figure 2: Non-uniform sampling

Let us focus on the case of two sequences (see figure 2). We denote by $l_a(t)$ the analog signal which is sampled. It is assumed to have a bandwidth limited to $1/T$. A sequence of samples y_1 taken at locations $nT + T_1$ is defined by :

$$y_1(t) = \sum_{n=-\infty}^{\infty} l_a(nT + T_1)\delta(t - nT - T_1) \tag{1}$$

and has a spectrum related to that of the analog signal by :

$$Y_1(\omega) = \frac{1}{T} \sum_{n=-\infty}^{\infty} \exp^{-2\pi j n T_1/T} L_a(\omega - 2\pi n/T) \tag{2}$$

A typical situation of spectrum repetition is shown in figure 3. It appears that when the spectrum is bandlimited to $1/T$ and the sampling rate is $1/T$, only two repeated versions of the initial spectrum interfere at the same time for any value of the frequency. It thus means that such a sequence of samples provides, for instance from 0 to $1/T$, a linear combination of the initial spectrum with two complex weights which are known when the sampling phase is known. A second sequence y_2 with another sampling phase T_2 would provide another linear combination :

$$Y_2(\omega) = \frac{1}{T} \sum_{n=-\infty}^{\infty} \exp^{-2\pi j n T_2/T} L_a(\omega - 2\pi n/T) \tag{3}$$

Considering frequences between 0 and 1/T, we can rewrite equations 2 and 3 as follows :

$$Y_1^+(\omega) = \frac{1}{T}\left[L_a^+(\omega) + e^{\frac{-2\pi j T_1}{T}} L_a^-(\omega - \frac{2\pi}{T})\right] \tag{4}$$

$$Y_2^+(\omega) = \frac{1}{T}\left[L_a^+(\omega) + e^{\frac{-2\pi j T_2}{T}} L_a^-(\omega - \frac{2\pi}{T})\right] \tag{5}$$

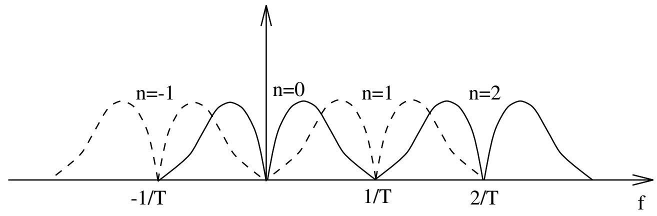

Figure 3: Repetition of the spectra around n/T

were $L_a^-(\omega)$ and $L_a^+(\omega)$ represent respectively the negative and positive parts of the baseband spectrum (n=0 – see figure 3). This makes a linear system with L_a^- and L_a^+ as unknowns and means that by solving these two equations, the spectrum of the analog input signal can be recovered from 0 to $1/T$. The same result holds from $-1/T$ to 0. Consequently, we are able to completely recover the input signal $L_a(\omega)$. As a consequence, $l_a(t)$ is recovered by taking the inverse Fourier transform of $L_a(\omega)$:

$$\begin{aligned} l_a(t) &= \sum_{k_1=-\infty}^{\infty} y_1(k_1T+T_1)\,\frac{\sin\left[\pi(t-k_1T-T_2)/T\right]}{\sin\left[\pi(T_1-T_2)/T\right]} \\ &\times \ \operatorname{sinc}\left[\pi(t-k_1T-T_1)/T\right] \\ &- \sum_{k_2=-\infty}^{\infty} y_2(k_2T+T_2)\,\frac{\sin\left[\pi(t-k_2T-T_1)/T\right]}{\sin\left[\pi(T_1-T_2)/T\right]} \\ &\times \ \operatorname{sinc}\left[\pi(t-k_2T-T_2)/T\right] \end{aligned} \tag{6}$$

This formalism is relevant for the vertical description of interlaced images. As a matter of fact, if the vertical sampling distance is denoted h within a frame, the theoretical range of the vertical resolution is upper limited by $1/2h$. In each field, there is a 2-fold vertical downsampling compared to the vertical frame sampling rate. As a consequence, each field contains aliasing. Nevertheless, considering the assumptions made previously about the motion, two successive fields provide two vertical sets of points which, in general, correspond to different vertical sampling phases. Therefore, by adequately combining the vertical information of two successive fields, the analog information and the associated vertical resolution "hidden behind these two fields" can theoretically be recovered exactly. In particular, the complete vertical resolution can be recovered.

3 Motion Estimation Using the General Sampling Theorem

In the HAMLET work, the motion of any field is estimated between this field and the two previous ones. We make the assumption that the motion from field 1 to field 3 is uniform. Actually, it is only required that the images can be partitioned into areas with a translational motion. We nevertheless make the calculations for the whole picture, but for the vertical direction only. The problem is illustrated by figure 4. We assume that fields 1 and 3 provide samples of their associated luminance signals in vertical positions $2k_1h - h$ and $2k_3h - h$ respectively, and that field 2 provides samples located in $2k_2h$, where the k_i are all integers. If y_0 is assumed to be the correct motion vector, it means that we are able to compute a prediction of the lines of field 3 from the information contained in fields 1 and 2. If the assumption of a uniform motion holds, the lines $2k_1h - h$ of field 1 provide samples located in $2k_1h-h+2y_0$ in field 3. Similarly, the lines $2k_2h$ of field 2 provide samples located in $2k_2h + y_0$ in field 3. If we assume that the analog luminance is vertically bandlimited to $1/2h$, we know 2 sequences of samples for field 3 and we are able to recover the exact analog signal associated with this field. This is done by applying formula 6 with $T = 2h$, $T_1 = 2y_0 - h$ and $T_2 = y_0$. This analog signal taken at locations $2k_3h - h$ provides an estimate of the exact lines of field 3 from the information contained in fields 1 and 2. The motion estimation procedure will provide the value of y_0 which minimizes a distance between the known luminance samples of field 3 and their associated estimates computed from fields 1 and 2. Any classical method can be used. For instance, block matching with exhaustive search or gradient-based approaches can be considered.

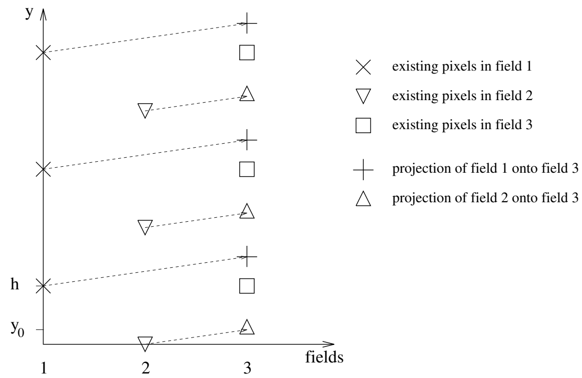

Figure 4: Motion estimation of field 3

4 Conversions between Interlaced and Progressive Scanning Formats

This section deals with conversions between progressive and interlaced scanning formats. As it refers to the coder side (see figure 1), an interlaced to progressive conversion may use some high-cost techniques. This is fortunate since low-cost linear deinterlacing (e.g. spatial and temporal linear interpolation) does not offer a sufficient picture quality at the input of the coder in order to improve the coding efficiency.

Motion compensated *deinterlacing* (25Hz-interlaced to 50Hz-progressive conversion) is really similar to motion estimation. Instead of estimating existing lines, one has to compute missing points. The estimation of the missing lines can be made from the previous field and the current one as represented in figure 5 : the generalized sampling theorem can be applied to the existing pixels of field 3 and the projections of field 2 in order to reconstruct the analog signal hidden behind field 3 and then resample this signal at the points of the missing pixels. Of course, this analog signal has not to be explicitly calculated. All the above mentioned operations can be performed in the digital domain. Detailled interpolation formulas can be found in [5].

In the situation that corresponds to a motion vector which is an odd multiple of the distance between two lines within a frame, namely h, the two sequences are produced with the same sampling phase. It is not possible to recover the exact analog signal anymore. In this situation, a fallback mode has to be defined. We decided to compute the missing point by averaging between the two surrounding lines in the current field.

Interlacing (50Hz-progressive to 25Hz-interlaced conversion) can easily be achieved by filtering the progressive source through a vertical low-pass filter and discarding the useless lines in order to generate the interlaced fields. The low-pass filtering is used to reduce the vertical definition according to the Kell factor. This factor is meant to reduce the line flicker that appears on bright and sharp horizontal edges when displayed in an interlaced format [1]. The impulse response of such filter (known as "HHI filter") is listed below :

-4	8	25	-123	230	728	230	-123	25	8	-4	/1000

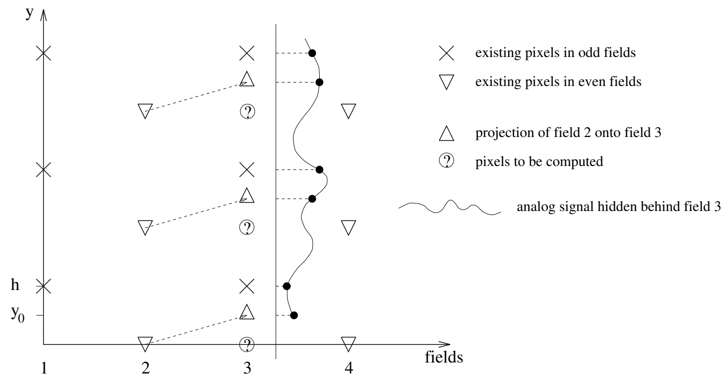

Figure 5: Deinterlacing the third field

As represented in figure 1, other interlaced/progressive conversions (25Hz-i to 25Hz-p and 25Hz-p to 25Hz-i) may be seen as a cascade of two modules, already or further described in this work.

5 Conversions Between Different Spatial Formats

In this section, we will focus on conversions between different spatial resolutions, as needed for conversions between High-Definition and Standard-Definition television formats. Again, a distinction has to be made between the two different scanning formats.

When working with a *progressive format*, spatial conversions are easily processed by using a chain of down-/up-sampling and low-pass filtering. As the scaling factor may be fractional, a digital *down-conversion* requires the use of both up- and down-samplers, as illustrated in figure 6 in the particular case of a 3/2 down-conversion. The low-pass filter is used to avoid the aliasing effect and suppresses the frequencies that could fold back to the low spectrum during the down-sampling operation. As these operations are not time invariant – due to the up-/down-sampling operators – they cannot be written in terms of a simple filtering expression but as a filter whose impulse response periodically depends on the output samples to be calculated [7]. An *up-conversion* also requires the use of both up- and down-samplers. Again, the up-converted sequence may be expressed with a filter whose impulse response depends on the output samples to be calculated.

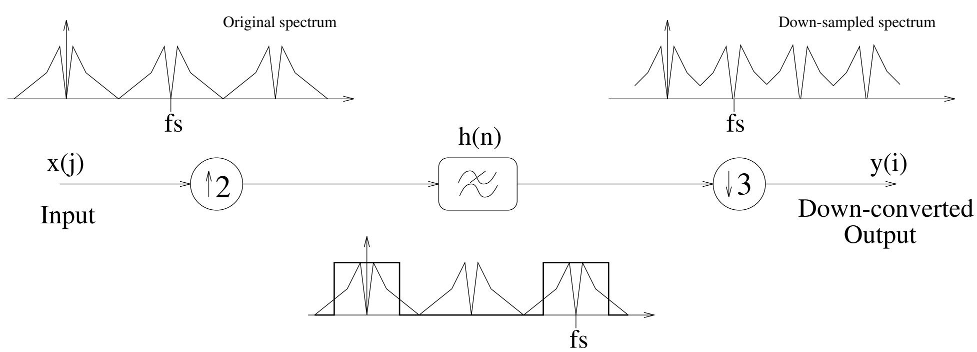

Figure 6: Down-sampling by a factor of 3/2

Concerning spatial conversions between *interlaced formats*, a first solution would be to merge the interlaced fields into progressive frames and apply the same processing as in the previous subsection.

Unfortunately, this technique produces ghost-effects and motion judder in moving parts of the picture. Another way would be to perform intra-field conversions. However, it does not take into account the aliasing present in each field. In order to get rid of this intra-field aliasing, an improved method can be based on the general sampling theory : the interlaced image is first deinterlaced (as explained in section 2) and its progressive version can be further up- or down-converted. As last step, reinterlacing has to be performed. Mixing all these operations together leads to an overall digital and motion-based method for interlaced image. Complete equations are developed in [8].

6 Temporal Conversions

Temporal down- and up-conversions revert basically to the same problem as for spatial down- and up-sampling, but for the temporal dimension. Again, the same basic outline as represented in figure 6 can be used. However, this problem slightly differs from the previous one due to the difficulty of implementing temporal filters and the presence of the judder effect likely to damage the source in case of inadequate temporal filtering. Moreover, the judder effect occurs when the difference between two adjoining frames is too high for the eye to reconstruct the missing movement existing between these images. This effect depends on several parameters : the frame rate, camera integration time, viewing distance, screen/picture resolution, screen brightness and picture contents [7]. These parameters are interacting: for example, we can reduce the exposure time (in order to improve the sharpness of the sequence) and still limit the judder effect by simultaneously increasing the shooting (and display) frequency.

By taking care of performing a low-pass temporal pre-filtering (e.g. bilinear filtering), it is possible to *down-convert* a sequence by simply throwing images away. However, in order to achieve good results (absence of judder), the exposure time of the original sequence must be large enough in order to avoid multiple contours as result of the low-pass (digital) pre-filtering. Ideally, this exposure time should cover the elapsed time between two successive images. Also, the low-pass pre-filtering must have enough taps in order to render a motion blur that covers the elapsed time between two successive images of the down-converted sequence. However, a longer filter implies an larger frame memory in order to perform the temporal filtering. It may run counter to the low-cost requirement considered here. This can be solved by using an appropriate exposure time at the camera. With respect to these, the judder appearing in quick moving parts of the scene can be avoided but to the detriment of some increased blur in these same areas. Concerning low-cost *up-conversions*, if the exposure time of the input sequence is high enough, it can simply be performed by repeating existing frames. This solution is used for displaying films on television.

When the shooting conditions are such as it is not possible to use some low-cost techniques or when there is a need to offer an improved quality and avoid additional blur in quick motion areas (broadcasting side), motion compensated conversions have to be used. For conversions implying *progressive formats*, a usual interpolation techniques may be used. For example, the pixels of the missing frames can be obtained by a temporal interpolation along the motion direction. Since the interpolated fields have to be inserted between those of the original sequence (e.g. 50Hz to 60Hz frame rate conversion), an assignment problem may occur since the blocks, once translated along the attached movement, do not necessary create a complete partition in the interpolated field. This problem can be solved by extending the matching area outside the elementary segmentation blocks in order to ensure at least the full covering of the inserted field. Uncovered areas will then disappear by giving rise to growing overlapped areas. For these areas, a choice has to be made between several block candidates. The best matching one can be selected by the use an absolute minimum absolute difference criteria [9]. Furthermore, a simple interpolation along the direction of the motion does not treat uncovered regions correctly. In such a case, it is better to leave the interpolation technique and not consider information from both frames surrounding the frame to be interpolated anymore, but consider information from the future only.

For the *interlaced format*, a deinterlacer has to be used prior to the temporal interpolator and the up-converted sequence has then to be reinterlaced. Again, using an algorithm that makes use of the general sampling theory improves the deinterlacing step and the up-conversion by the same way.

7 Conclusion

This paper specified a deinterlacer that makes use of the general sampling theory. This theory handles the aliasing present in each field of an interlaced sequence and allows a perfect reconstruction of the progressive frame hidden behind those fields. Such deinterlacer has been used within the HAMLET/WP2 in order to study the coding efficiency of deinterlaced (progressive) sequences compared to their interlaced sources. Other conversions, spatial as well as temporal, were also considered here and further references about the generalized sampling theory have been made in order to improve these conversions when interlaced formats were involved.

References

[1] S. Pigeon and P. Guillotel, "Advantages and Drawbacks of Interlaced and Progressive Scanning Formats", *CEC RACE/HAMLET Deliverable n° R2110/WP2/DS/R/004/b1*, June 1995.

[2] P. Guillotel and S. Pigeon, "Coding Efficiency Comparisons of both Interlaced and Progressive Scanning Formats", *CEC RACE/HAMLET Deliverable n° R2110/WP2/DS/S/012/b1*, December 1995.

[3] J. Pandel, "Introduction and Overview about the TRANSIT Project", *European Workshop and Exhibition on Image Format Conversion and Transcoding*, Berlin, Germany, March 1995.

[4] P. Delogne, L. Cuvelier, B. Maison, B. Van Caillie and L. Vandendorpe, "Improved Interpolation, Motion Estimation and Compensation for Interlaced Pictures", *IEEE Transactions on Image Processing*, Vol. 3, No. 5, September 1994, pp. 482-491.

[5] L. Vandendorpe, L. Cuvelier, B. Maison, P. Queluz and P. Delogne, "Motion Compensated Conversion from Interlaced to Progressive Formats", *Signal Processing : Image Communication*, Vol. 6, No. 3, June 1994, pp. 193-211.

[6] J. L. Yen, "On the nonuniform sampling of bandlimited signals", *IRE Trans. Circuit Theory*, Vol. CT-3, December 1956, pp. 251-257.

[7] S. Pigeon, L. Vandendorpe, L. Cuvelier and B. Maison, "Specification of a Generic Format Converter", *CEC RACE/HAMLET Deliverable n° R2110/WP2/DS/S/006/b1*, September 1995.

[8] L. Vandendorpe, L. Cuvelier, B. Maison and P. Delogne, "Improved Motion-compensated Conversion between Interlaced TV and HDTV formats", *International Workshop On HDTV'93*, Ottawa, Canada, October 1993.

[9] L. Vandendorpe, P. Delogne and D. Boucquey, "Generalized interpolators for advanced movement-compensated 50Hz-60Hz conversion of interlaced sequences", *1995 IEEE International Conference on Image Processing*, Vol. 2, pp. 237-240.