Motion-compensated interpolation for face-centered -orthorhombic sampled video sequence

Ning Guan (关 宁)*, Xu Zhang (张 旭), and Hongda Chen (陈弘达)

State Key Laboratory on Integrated Optoelectronics, Institute of Semiconductors,
Chinese Academy of Science, Beijing 100083, China

**Corresponding author: guanning@semi.ac.cn*

Received November 22, 2011; accepted December 28, 2011; posted online February 24, 2012

Face-centered orthorhombic (FCO) sampling can be implemented more easily on CMOS image sensors than on other video acquisition devices. The sampling efficiency of FCO is the highest among all three-dimensional (3D) sampling schemes. However, interpolation of FCO-sampled data is inevitable in bridging human perception and machine-vision algorithms. In this letter, the concept of motion compensation is borrowed from deinterlacing, which displays interlaced videos on progressively scanned devices. The combination of motion estimation based on intrafield interpolated frames and motion-compensated interfield interpolation is found to provide the best performance by evaluating different combinations of motion estimation and interpolation.

OCIS codes: 100.3010, 110.3010, 110.4155, 100.2000.
doi: 10.3788/COL201210.061001.

The pixel rate, which is defined as the number of pixel values that can be extracted from the pixel array during a given period, is limited in resource-constraint situations, which include high-speed and ultra-low-power video acquisition. The efficiency of the sampling scheme is critical in increasing the overall performance. The multidimensional sampling theory[1] proves that face-centered orthorhombic (FCO) sampling (as shown in Fig. 1(b)) offers the highest sampling efficiency among all three-dimensional (3D) sampling schemes just as hexagonal pixels (Fig. 1(a)) do for two-dimensional (2D) image sampling.

Conventional video sampling apparatuses such as vacuum tubes and charge-coupled devices (CCDs) are restricted by their scanning readout nature and are barely capable of realizing interlaced scanning. When complementary metal oxide semiconductor (CMOS) image sensors (CISs) are realized, FCO sampling[2] can be implemented because the readout circuitry of CISs resembles random-access memories.

The FCO-sampled sequences have to be interpolated up to full resolution to bridge FCO-sampled video sequences and existing video-processing algorithms. This problem is similar to the deinterlacing[3] process, which displays interlaced video sequences on progressively scanned screens. However, the final judgment on deinterlacing is only human perception. FCO interpolation must further fulfill the need for machine-vision applications. Thus, its absolute accuracy is also of main concern and not just human feelings.

Existing deinterlacing algorithms are categorized into linear, motion-adaptive, and motion-compensated methods. Linear algorithms include intra- and interfield interpolations. Some linear algorithms[4] even develop interpolating coefficients based on 3D sampling as was also done by Guan *et al.*[2] Motion-adaptive deinterlacing algorithms[5−7] use motion detectors to switch between intra- and intrafield interpolating methods in different areas of a field. They are practical for display because human eyes are less sensitive to the details of moving objects. However, they are improper for general-purpose interpolation because no extra information is added where intrafield interpolation is used.

On the contrary, motion-compensated algorithms[8−10] add extra information over the whole field. Generally speaking, these algorithms first shift corresponding areas in the former and latter fields using a motion vector (MV). Then, they interpolate based on the current and shifted fields as if they represent the same stationary scene. However, the motion estimators (MEs) and interpolators cannot be directly used due to the different sampling schemes—one is the interlaced scheme and the other is FCO.

Therefore, we aim to determine which combination of ME and motion-compensated interpolator is most suitable for reconstructing FCO-sampled videos. Both objective and subjective criteria are used in the evaluation.

The original signal on the focal plane is a time-varying 2D illumination $\psi(x, y, t)$, which is a 3D signal if the magnitudes of x, y and t are ignored. The multidimensional Nyquist condition is not simply the superposition of multiple one-dimensional (1D) criteria[11].

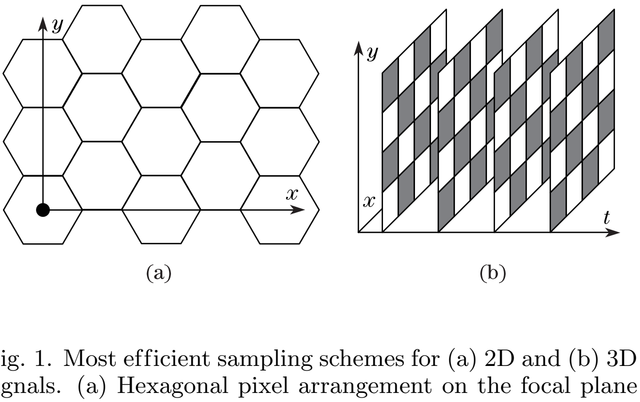

Fig. 1. Most efficient sampling schemes for (a) 2D and (b) 3D signals. (a) Hexagonal pixel arrangement on the focal plane of the (b) FCO sampling scheme, in which only white pixels are read out at each sampling time.

Thus, the sampling scheme can be changed while still fulfilling the Nyquist condition. The multidimensional sampling theory[1] proves the efficiency of the sampling scheme in Fig. 1(b), which is called FCO sampling. This scheme provides the highest sampling efficiency among all 3D sampling schemes.

With conventional line-scanning tubes and CCDs, FCO sampling is hard to implement. However, with the dawn of the CIS era, FCO sampling with CIS technology has become quite easily realizable. As shown in Fig. 2, adjacent pixels have separate resets (RST0 to RST3) and shutters (SHT0 to SHT3) while still sharing column-wise output buses. By repeating this structure over the whole array and connecting control pins of corresponding pixels together, a pixel array of eight control pins (four RSTs and four SHTs) is obtained instead of the conventional two pins (RST and SHT). Such structure can be accomplished using multiple layers of metals[2]. The emerging back-side illumination[12] could minimize the side effects of more metal layers.

As shown in the upper two lanes in Fig. 3, initially, both RST0, 3 and SHT0, 3 are on. Pixels 0 and 3 are in the reset (rst in Fig. 3) stage. When RST0, 3 and SHT0, 3 are turned off, pixels 0 and 3 start to be exposed (exp in Fig. 3). At the end of the exposure, SHT0, 3 is briefly turned on for the transfer of photocarriers to the capacitor (floating diffusion). These carriers are then held on the capacitor and can be read out (rd in Fig. 3) through column buses. Considering that control pins are separated, adjacent pixels in different stages of operation can be obtained. Pixels 0 and 3 are in the exposure stage, whereas pixels 1 and 2 are in the readout and then reset stages, as shown in the lower two lanes in Fig. 3. Thus, FCO sampling is realized in this kind of CIS.

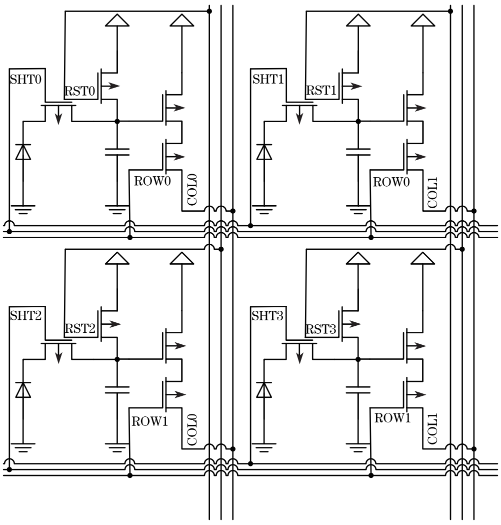

Fig. 2. Schematic of the control-pin-separated pixels which are capable of realizing FCO sampling.

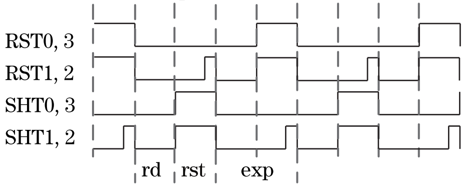

Fig. 3. Timing diagram of the control-pin-separated pixels, which realize a typical FCO sampling scheme. The italic abbreviations above and below the signal lanes indicate different stages of pixels. “rst”, “exp”, and “rd” stand for reset, exposure, and readout, respectively.

The first step to realizing a motion-compensated interpolator is to estimate MVs based on the FCO-sampled video sequence. The performance of motion-compensated interpolation largely depends on the accuracy of motion estimation. A different motion estimating algorithm for FCO-sampled video sequences is evaluated to develop a proper motion-compensated interpolating method.

Block-based motion estimation is widely used in deinterlacing applications because of its simplicity and neutrality. It introduces no extra steps to mesh the scene or identify texture and provides promising results with further MV correction.

The sum of absolute difference (SAD) is mused as the matching criterion in the following MEs. However, calculating SAD between two consecutive fields is improper because they are of different polarities. Hence, a solution to this problem is needed.

The most straightforward way is to reduce the FCO-sampled video sequence to full resolution with intrafield interpolation, which is accomplished by averaging the neighboring pixels

$$\begin{aligned}\psi_{n,\,\mathrm{intra}}(x,\,y)=&[\psi_{n,\,\mathrm{FCO}}(x-1,\,y)+\psi_{n,\,\mathrm{FCO}}(x+1,\,y)\\&+\psi_{n,\,\mathrm{FCO}}(x,\,y-1)\\&+\psi_{n,\,\mathrm{FCO}}(x,\,y+1)]/4,\end{aligned}\tag{1}$$

if (x, y) is not sampled in the current field. Therefore, the problem is the motion estimation for a full-resolution video sequence $\psi_n(x,\,y)$, which calculates

$$\begin{aligned}\mathrm{SAD}_{\mathrm{block}}(d_x,\,d_y)=\sum_{x,y\in B_{\mathrm{block}}}&|\psi_{n-1}(x+d_x,\,y+d_y)\\&-\psi_n(x,\,y)|,\end{aligned}\tag{2}$$

and minimizes the $\mathrm{SAD}_{\mathrm{block}}$ over the searching region

$$\mathbf{d}_{m\text{-intra}}=\arg\left[\min_{(d_x,\,d_y)\in\mathrm{SR}}\mathrm{SAD}_{\mathrm{block}}(d_x,\,d_y)\right],\tag{3}$$

where SR represents the searching region, that is, the set of vectors that $\mathbf{d}_m$ can be.

Another option is to adopt the spatiotemporal interpolation[2] that requires no a priori knowledge upon the time-varying illumination on the focal plane

$$\begin{aligned}\psi_{n,\mathrm{sinc}}(x,\,y)=&\frac{\pi^3}{8\pi+16}\left[\frac{16}{\pi^3}\psi_{n+1,\,\mathrm{FCO}}(x,\,y)\right.\\&+\frac{16}{\pi^3}\psi_{n-1,\,\mathrm{FCO}}(x,\,y)\\&+\frac{2}{\pi^2}\psi_{n,\,\mathrm{FCO}}(x-1,\,y)\\&+\frac{2}{\pi^2}\psi_{n,\,\mathrm{FCO}}(x+1,\,y)\\&+\frac{2}{\pi^2}\psi_{n,\,\mathrm{FCO}}(x,\,y-1)\\&\left.+\frac{2}{\pi^2}\psi_{n,\,\mathrm{FCO}}(x,\,y+1)\right],\end{aligned}\tag{4}$$

and to minimize the $\text{SAD}_{\text{block}}$ in Eq. (2) over the searching region such as that in Eq. (3). It is denoted as sinc interpolation, because it is based on the 3D sinc function of the FCO sampling[2]. Moreover, the MVs acquired through sinc-interpolated frames are called $\mathbf{d}_{m\text{-sinc}}$.

The polarity problem is also resolved under the assumption that the motion is extendable among successive n fields. If $n = 3$, the two fields before and after the current field are of the same parity. With these two fields, the candidate blocks can be compared, which should be symmetrically located with respect to the current block

$$\text{SAD}_{\text{bidir}}(d_x,\, d_y) = \sum_{x,y\in B_{\text{FCO}}} \big|\psi_{n-1,\,\text{FCO}}(x+d_x,\, y+d_y) - \psi_{n+1,\,\text{FCO}}(x-d_x,\, y-d_y)\big|. \quad (5)$$

Moreover, the MV is acquired through

$$\mathbf{d}_{m\text{-bidir}} = \arg\Big[\min_{(d_x,\,d_y)\in\text{SR}} \text{SAD}_{\text{bidir}}(d_x,\, d_y)\Big]. \quad (6)$$

Under the same assumption, the four-field motion estimation was proposed by Chang *et al.*[13]. It defines another SAD

$$\text{SAD}_{\text{field4}}(d_x,\, d_y) = \sum_{x,y\in B_{\text{FCO}}} |\psi_{n-2,\,\text{FCO}}(x+2d_x,\, y+2d_y) - \psi_{n,\,\text{FCO}}(x,\, y)|, \quad (7)$$

and minimizes the sum of two SADs

$$\mathbf{d}_{m\text{-field4}} = \arg\Big\{\min_{(d_x,\,d_y)\in\text{SR}} [\text{SAD}_{\text{bidir}}(d_x,\, d_y) + \text{SAD}_{\text{field4}}(d_x,\, d_y)]\Big\}. \quad (8)$$

Strictly speaking, an interpolator should achieve arbitrary subpixel accuracy, and subpixel MV is necessary. However, the interpolator for FCO sampling just serves as a bridge to existing image processing algorithms. If the one-pixel interpolator is accurate, an existing full-resolution interpolator can be used to achieve arbitrary subpixel accuracy.

The one-pixel accuracy can reduce the searching region by half. If the shifted and current blocks are of the same parity, no extra information is provided for the interpolation. Hence, an even criterion is applied for the MV

$$\text{mod}\,(d_x + d_y,\, 2) = 0, \quad (9)$$

which makes the parities of the shifted and current blocks different.

When MVs are available, most interpolating algorithms can be motion-compensated. Considering the restricted MVs in Eq. (9), the MVs will always point at existing pixels in adjacent fields. The interpolation methods based on these existing pixels can forbid the propagation of interpolating errors. Thus, these pixels can be reduced to shorter lists, which are backward projection, interfield interpolation, and spatiotemporal interpolation, all of which are motion-compensated.

Backward projection only finds missing pixels in the last motion-compensated field

$$\psi_{n,\,\text{mcproj}}(x,\, y) = \psi_{n-1,\,\text{FCO}}(x+d_x,\, y+d_y), \quad (10)$$

where $\psi(x, y)$ is missing. This method is denoted as "mcproj".

Motion-compensated interfield interpolation is described as

$$\psi_{n,\,\text{mcinter}}(x,\, y) = \frac{1}{2}\psi_{n-1,\,\text{FCO}}(x+d_x,\, y+d_y) + \frac{1}{2}\psi_{n+1,\,\text{FCO}}(x-d_x,\, y-d_y), \quad (11)$$

and denoted as "mcinter".

Spatiotemporal interpolation can also be motion-compensated:

$$\begin{aligned}\psi_{n,\,\text{mcsinc}}(x,\, y) = \frac{\pi^3}{8\pi+16}\Big[&\frac{16}{\pi^3}\psi_{n-1,\,\text{FCO}}(x+d_x,\, y+d_y)\\ &+ \frac{16}{\pi^3}\psi_{n+1,\,\text{FCO}}(x-d_x,\, y-d_y)\\ &+ \frac{2}{\pi^2}\psi_{n,\,\text{FCO}}(x-1,\, y)\\ &+ \frac{2}{\pi^2}\psi_{n,\,\text{FCO}}(x+1,\, y)\\ &+ \frac{2}{\pi^2}\psi_{n,\,\text{FCO}}(x,\, y-1)\\ &+ \frac{2}{\pi^2}\psi_{n,\,\text{FCO}}(x,\, y+1)\Big]. \quad (12)\end{aligned}$$

It is denoted as "mcsinc" just like the earlier two methods.

Unlike the evaluation of the deinterlacing process, subsampling standard test sequences[14] to FCO sample videos is improper because it introduces aliasing. As shown in Table 1, if the frame interval of conventional sampling is Δt, the no-aliasing f_t is $\sqrt{2}/2\Delta t^{-1}$, which is larger than $1/2\Delta t^{-1}$. However, the field interval between FCO subsampled videos is Δt. Therefore, the no-aliasing f_t is reduced to $\sqrt{2}/4\Delta t^{-1}$, which is smaller than $1/2\Delta t^{-1}$ and introduces aliasing. On the other hand, in a real FCO video acquisition device, the field interval of FCO sampling is $\Delta t/2$, keeping the pixel rate unchanged.

FCO sampling does not introduce spatial aliasing. Therefore, a temporal filter can be applied to alleviate the temporal aliasing effects. Considering the limited length of test sequences, an averaging filter is chosen instead of a high-order FIR filter:

$$\psi_{n,\,\text{ave}}(x,\, y) = \frac{1}{2}\psi_n(x,\, y) + \frac{1}{2}\psi_{n-1}(x,\, y). \quad (13)$$

Table 1. No-Aliasing Frequencies of Different Sampling Schemes

Sampling Scheme	No-aliasing		
	f_x^1	f_y^1	f_t^1
Conventional^2	1/2	1/2	1/2
Real FCO^2	1/2	1/2	$\sqrt{2}/2$
Subsampled FCO^3	1/2	1/2	$\sqrt{2}/4$

^{1}They are scaled by Δx^{-1}, Δy^{-1}, and Δt^{-1}, respectively. ^{2}The frame interval of conventional sampling is Δt, and the field interval of FCO sampling is $\Delta t/2$. Moreover, Δx and Δy are the same. ^{3}The field interval is Δt.

It generates at least 7 dB of attenuation to high-frequency components $f_t \in [\sqrt{2}/4\Delta t^{-1}, 1/2\Delta t^{-1}]$ and introduces no phase distortion.

As shown in Fig. 4, to evaluate an actual MC interpolator, the standard test video ψ_n is first filtered using a temporal low-pass (averaging) filter. The filtered result $\psi_{n,\,\mathrm{ave}}$ is then FCO subsampled and fed into a certain FCO ME to calculate the MV. A 16×16 (pixel) block and 30 × 30 pixel searching region. Thus, $d_x, d_y \in [-7, 7]$ and an exhaustive searching method is used to eliminate the effect of quick methods. This MV is then given to an MC interpolator, which produces $\psi_{n,\,\mathrm{interpolated}}$ ($\psi_{n,\,\mathrm{mcproj}}$, $\psi_{n,\,\mathrm{mcinter}}$, or $\psi_{n,\,\mathrm{mcsinc}}$). $\psi_{n,\,\mathrm{ave}}$ and $\psi_{n,\,\mathrm{interpolated}}$ are compared with the mean squared error (MSE) of a whole test sequence

$$\mathrm{MSE}=\frac{1}{P\times Q\times N}\sum_{n=1}^{N}\sum_{i=0}^{P-1}\sum_{j=0}^{Q-1} \cdot \left[\psi_{n,\,\mathrm{interpolated}}(i,j)-\psi_{n,\,\mathrm{ave}}(i,j)\right]^2. \qquad (14)$$

The results are listed in Table 2. The MSE of the no-MC interpolations are also calculated for comparison purposes.

Table 2 shows that the two motion-compensated interfield interpolators are suitable for reconstructing FCO-sampled videos. One (denoted as sinc-mcinter) uses the MV calculated from sinc interpolated frames ($\mathbf{d}_{m\text{-sinc}}$), and the other (denoted as intra-mcinter) uses MV calculated from intrafield interpolated frames ($\mathbf{d}_{m\text{-intra}}$). The slightly better performance of the sinc-mcinter is due to the better average accuracy of its ME.

$\psi_{n,\,\mathrm{ave}}$ is also motion-estimated by a block-search ME to calculate a reference MV. This reference MV is compared against the MV estimated by the FCO MEs. If the difference between these two MVs are (0, 0), the FCO ME is considered to provide the correct MV. As shown in Table 3, the average accuracy of $\mathbf{d}_{m\text{-sinc}}$ is higher than $\mathbf{d}_{m\text{-intra}}$.

However, the complexity of calculating a sinc-interpolated frame is higher than that of calculating an intrafield interpolated frame. Floating-point multiplication is required to realize sinc interpolation (4). Even if integer approximation is used, it still needs two terms

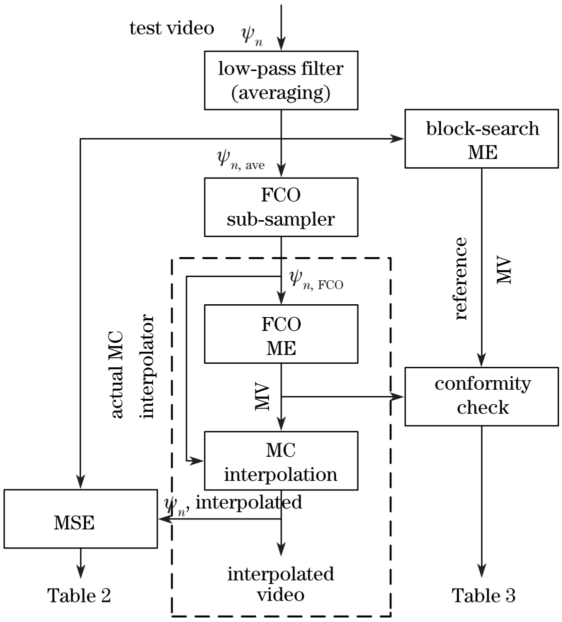

Fig. 4. Evaluation flow of the motion-compensated interpolator for the FCO-sampled video.

Table 2. MSE of Different MC Interpolation Methods

		CG*	CT*	FM*	HL*	MD*	NW*	SL*	Sum of MSEs
Intra		26.44	42.27	18.30	29.99	5.24	26.66	18.97	167.86
Number of MCs	Proj	59.89	3.79	62.82	7.43	4.34	15.32	13.90	167.49
	Inter	19.22	0.59	18.64	2.24	1.17	4.19	4.12	50.16
	Sinc	15.92	8.75	12.08	6.82	1.61	7.13	5.32	57.64
$\mathbf{d}_{m\text{-Intra}}$	Mcproj	23.70	3.72	25.61	6.01	2.71	9.59	7.70	79.04
	Mcinter	10.34	1.01	14.08	2.79	1.14	5.90	4.99	40.25
	Mcsinc	40.25	40.36	21.59	34.00	7.98	35.71	18.10	198.00
$\mathbf{d}_{m\text{-Sinc}}$	Mcproj	23.92	3.72	25.15	5.94	2.68	9.63	7.58	78.61
	Mcinter	9.54	0.60	13.48	2.14	0.98	5.30	4.34	36.37
	Mcsinc	39.88	40.05	21.02	33.58	7.88	35.42	17.72	195.55
$\mathbf{d}_{m\text{-Bidir}}$	Mcproj	28.94	3.89	34.62	6.65	3.25	14.50	10.84	102.68
	Mcinter	13.21	0.79	19.89	2.85	1.37	9.05	6.70	53.86
	Mcsinc	41.36	40.13	23.29	33.91	8.04	36.89	18.64	202.26
$\mathbf{d}_{m\text{-Field4}}$	Mcproj	27.18	3.84	30.63	6.27	3.00	11.85	9.17	91.95
	Mcinter	12.92	0.75	17.67	2.59	1.26	7.20	5.72	48.12
	Mcsinc	41.38	40.12	22.63	33.81	8.01	36.24	18.30	200.48

*Note: Abbreviation of test sequences. CG: coastguard; CT: container; FM: foreman; HL: hall; MD: mother–daughter; NW: news; SL: silent.

Table 3. Accuracy of Different Motion-Estimation Methods

	CG*	CT*	FM*	HL*	MD*	NW*	SL*	Average
	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)
$\mathbf{d}_{m\text{-intra}}$	94.93	97.93	84.49	97.26	89.90	97.03	92.25	93.40
$\mathbf{d}_{m\text{-sinc}}$	88.06	99.88	82.99	99.47	96.93	97.66	95.77	94.39
$\mathbf{d}_{m\text{-bidir}}$	86.08	96.94	62.63	95.20	80.20	93.51	89.44	86.29
$\mathbf{d}_{m\text{-field4}}$	88.97	97.54	70.61	97.06	84.53	94.20	91.63	89.22

more than that of intrafield interpolation (1).

Another disadvantage of both mcinter methods is that they introduce a one-field delay. This delay might be intolerable in several real-time applications.

A moving scene with a stepper-motor-driven linear stage and a poker card (two of spades) is created. The background has black and white fringes with 1-cm widths. A FCO-sampled video sequence of 64 fields is acquired using the control-pin-separated pixel array[2].

In the original interlaced frame (Fig. 5(a)), the spade nor the number "2" is unrecognizable. However, using the sinc function of the FCO, a frame (Fig. 5(b)) showing the spade can be reconstructed. However, "2" is still vague.

The intra-mcinter (Fig. 5(c)) and sinc-mcinter (Fig. 5(d)) methods both make the spade more concrete. In the interpolated frame (Fig. 5(c)) of intra-mcinter, a clearer "2" can be found. Although this phenomenon is subjective, it indicates that intra-mcinter might be better even though its MSE is larger than that of sinc-mcinter.

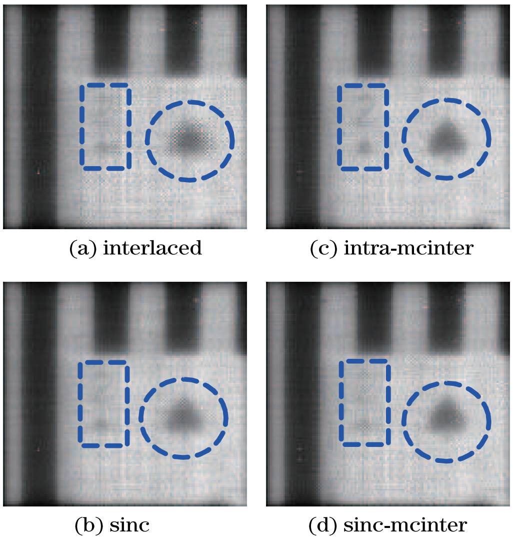
(a) interlaced
(c) intra-mcinter
(b) sinc
(d) sinc-mcinter

Fig. 5. (a) Original interlaced frame and interpolation frames of a real FCO sampled video, and the one using (b) the sinc function of FCO, (c) intra-mcinter, and (d) sinc-mcinter.

In conclusion, two suitable motion-compensated interpolators are determined by evaluating various combinations of MEs and interpolators with low-pass-filtered FCO subsampled standard test sequences. The interpolations of real FCO-sampled video sequences show that intra-mcinter is better, using MV calculated by intrafield interpolated frames to motion-compensate an interfield interpolation. This method could be the foundation of a more rigorous interpolating algorithm.

This work was supported by the National Natural Science Foundation of China (Nos. 60877035, 61076023, and 90820002) and the National "863" Program of China (Nos. 2011CB933203 and 2011CB933102).

References

1. D. P. Peterson and D. Middleton, Inf. Control **5,** 279 (1962).
2. N. Guan, X. Zhang, and B. Liu, Sci. China Inf. Sci. (to be published).
3. G. de Haan and E. B. Bellers, Proc. IEEE **86,** 1839 (1998).
4. G. Microchip, Inc. "Preliminary data sheet of Genesis gmVLD8, 8 bit digital video doubler, version 1.0", June 1996.
5. S. Lin, Y. Chang, and L. Chen, IEEE Trans. Consum. Electron. **49,** 1256 (2003).
6. D. Han, C. Y. Shin, S. J. Choi, and J. S. Park, IEEE Trans. Consum. Electron. **45,** 690 (1999).
7. V. Markandey, T. Clatanoff, R. Gove, and K. Ohara, IEEE Trans. Consum. Electron. **40,** 735 (1994).
8. Y.-L. Chang, C.-Y. Chen, S.-F. Lin, and L.-G. Chen, in *Proceedings of International Conference on Image Processing* **2,** III-693-6 (2003).
9. O. Kwon, K. Sohn, and C. Lee, IEEE Trans. Consum. Electron. **49,** 198 (2003).
10. K. Sugiyama and H. Nakamura, IEEE Trans. Consum. Electron. **45,** 611 (2000).
11. E. Dubois, Proc. IEEE **73,** 502 (1984).
12. H. Wakabayashi, K. Yamaguchi, M. Okano, S. Kuramochi, O. Kumagai, S. Sakane, M. Ito, M. Hatano, M. Kikuchi, Y. Yamagata, T. Shikanai, K. Koseki, K. Mabuchi, Y. Maruyama, K. Akiyama, E. Miyata, T. Honda, M. Ohashi, and T. Nomoto, in *Proceedings of Solid-State Circuits Conference Digest of Technical* 410 (2010).
13. Y. Chang, S. Lin, C. Chen, and L. Chen, IEEE Trans. Circuits Syst. Video Technol. **15,** 1569 (2005).
14. "YUV video sequences", http://trace.eas.asu.edu/yuv/.