

A system approach to high-quality picture-rate conversion

Chris Bartels
Claus Nico Cordes
Bram Riemens
Gerard de Haan

Abstract — Various techniques have been implemented to improve the motion portrayal of flat-panel displays, of which the widespread introduction of motion-compensated picture-rate conversion systems is an essential part. However, a careful design of such systems is critical as they have the potential to introduce perceptually annoying artifacts. In this paper, an overview of the required “ingredients” of high-quality motion-compensated picture-rate conversion is presented, as implemented in state-of-the-art systems-on-chip. The visual impact of individual components on picture quality is illustrated.

Keywords — Motion blur, motion judder, motion compensation, picture-rate conversion, frame-rate conversion.

DOI # 10.1889/SID18.11.922

1 Introduction

Motion portrayal is an important aspect of picture quality. Motion blur affects the “dynamic” resolution of a display and in this respect, flat-panel-display technologies such as LCD technology have long remained inferior to CRT displays.¹ The two characteristics that caused this were (1) low-response times and (2) the picture sample-and-hold effect.

Over time, the slow response has been addressed by panel makers through material innovations, while the remaining effect has been further reduced by display processing such as overdrive.²

The negative consequences of the picture sample-and-hold have been targeted by increasing the display picture rate from 50/60 Hz to 100/120 Hz, and even to 200/240 Hz. However, since video content is mostly captured at relatively low picture rates, such as 24 Hz for film or 50/60 Hz for video cameras, high-quality picture-rate conversion (PRC) is now essential to profit from fast-responding panels.

This poses a significant challenge, as the eye-tracking of the human viewer requires PRC algorithm designers to apply motion estimation (ME) and motion compensation (MC) techniques.^{3,4} Simpler PRC methods using repetition or linear interpolation of images cause motion blur and/or judder which render the advances in display technology useless. Motion-compensated PRC (MC-PRC) can enable perfect motion portrayal, but, to this day, there is no design without occasional failure. Limiting the occurrence frequency of failures and avoiding perceptually annoying artifacts is challenging; particularly now, recent high-quality displays make even minor artifacts visible.

In this paper, we shall discuss the key requirements and lessons learned in the design of state-of-the-art MC-PRC systems, such as those discussed in Ref. 6.

In Sec. 2, we show the components of a high-quality PRC system. In Secs. 3–7, these will be illustrated individually by showing their visual impact from an end-user perspective. We draw our conclusions in Sec. 8.

2 Components of MC-PRC

MC-PRC is based on the estimation of the displacement path or “motion trajectory” of all pixels between consecutive frames, *i.e.*, (dense) motion estimation. Early MC-PRC algorithms relied on relatively simple interpolation technologies to profit from the ME. However, since the introduction of MC-PRC in CRT-TV systems,⁵ display technology has improved considerably, modern displays exhibit high spatial resolutions, high contrast ratios, and large screen sizes. As a consequence, the quality requirements for MC-PRC have increased over time, as the visibility of even small processing artifacts on these improved displays is likely.

This has led to more sophisticated designs of the MC-PRC system (illustrated in Fig. 1), where we now broadly recognize the following components: robust film mode detection, true motion estimation, robust interpolation with occlusion detection, and global fallback.⁶

In the following sections, the components are described one by one.

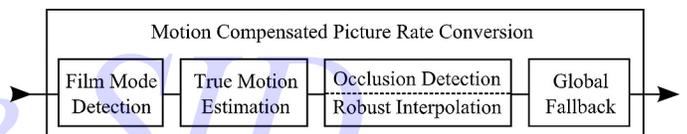


FIGURE 1 — Processing “chain” of a picture-rate-conversion system.

^aIn this paper we consider progressively scanned video. A practical system also includes de-interlacing, which we consider outside the scope of this contribution.

Received 04/12/10; accepted 07/13/10.

C. Bartels is with Eindhoven University of Technology, Department of Electrical Engineering, Den Dolech 2, 5600MB Eindhoven, The Netherlands; e-mail: chris@ics.ele.tue.nl.

C. N. Cordes and B. Riemens are with NPX Semiconductors Research, Eindhoven, The Netherlands.

G. de Haan is with Philips Research Laboratories, Eindhoven, The Netherlands.

© Copyright 2010 Society for Information Display 1071-0922/10/1811-0922\$1.00.

3 Robust film mode detection

Most display devices and broadcasters process standardized 50/60-Hz video signals. Content with a different picture rate, *e.g.*, 24/25-Hz film, is typically converted before transmission by means of picture repetition.^b As a result, video signals entering the TV may originate from a source with a different picture rate than that of the transmitted signal. The task of the film mode detection is to identify the original picture rate and remove repeated video frames. In the past, this was particularly relevant for film content, which was converted from 25 to 50 Hz using 2:2 pull-down or from 24 to 60 Hz using 3:2 pull-down. Nowadays, due to a wider variety of input formats, more of these pull-down patterns can be observed in 50/60-Hz video signals, such as:

- 2:3:3:2 or 2:2:2:4 for 24–60 Hz (*e.g.*, DV-Cam)
- 2:2:3:2:3 for 25–60 Hz
- 4:4 for 15–60 Hz (*e.g.*, YouTube or animation)
- 2:2:2:2:2:2:2:2:2:3 for 24–50 Hz (“TelecineB”)

Obviously, the film mode detection must also identify input signals in which no repetition pattern occur, typical for news-reports or sports-events broadcasts.

The film mode detection has to correctly identify these patterns in order to enable MC-PRC for such content. It identifies which input fields originate from a single film image and determines the time instances of the interpolated output images; *e.g.*, for 2:2 pull-down with 25-Hz film content and a 100-Hz output, time instances of 1/4, 1/2, and 3/4 of the time difference between successive input pictures are used. In case of 3:2 pull-down with 24-Hz film content for 120-Hz output, time instances of 1/5, 2/5, 3/5, and 4/5 are used. The film mode detection controls which images are sent to the rest of the MC-PRC system: an incorrect detection typically results in (remaining) motion judder.

Despite the large influence on the conversion result, there are few scientific publications on the topic of film mode detection. In publications on PRC, the film mode detection is sometimes described as part of the system, but rarely elaborated upon, *e.g.*, see Refs. 7 and 8. In most cases the detection is limited to 3:2 pull-down.⁹ Industrial relevance of the topic can be recognized from patent literature, *e.g.*, see Refs. 10–14.

The most common strategy for film mode detection is based on an image difference measure between successive input fields.^{9–14} Alternatively, a motion measure between successive input fields can be used.^{8,15} Once the measure is determined, a sequence of such values is analyzed and checked against expected patterns according to a set of supported film modes. A comparison between field N and $N - 2$ is particularly suitable for 3:2 pull-down detection. This frame difference will be low for one out of five input field pairs, which yield the same interlace phase. For detection of 2:2 film mode or any repetition pattern of just two

^bAlso referred to as $n:m:\dots$ pull-down, with n, m, \dots denoting the repetition factors of subsequent original video frames.

fields, comparison between two consecutive input fields is required. The main challenges with image-comparison-based methods we recognize as:

Motion: When there is a large amount of motion in the scene, differences in consecutive fields are apparent, and reliable film mode detection is easily achieved. This is not the case for scenes with “little” motion, *i.e.*, only a small part of the image is moving, or “small” motion, *i.e.*, the velocity of the motion is very low. For the former, a field difference can only be detected based on a small fraction of the available pixels in the image. For the latter, image differences are small even in the presence of motion. Note that both cases occur very often in almost any content and failure to detect the correct film mode leads to annoying artifacts, even with limited motion. This results in challenging requirements on the sensitivity of the field-difference detector.

Noise: Analog noise results in differences between fields that originate from the same film image but without any motion involved. Signals that originate from PAL or NTSC coding may suffer from a color subcarrier residual signal.

Digital “noise” is caused by artifacts in the transmission chain due to, *e.g.*, MPEG-2 or H.264 coding. This may include quantization errors, ringing, blockiness, and mosquito noise. Note that video encoders often use different quality levels depending on the picture coding type in a GOP (group of pictures); *e.g.*, a relatively strong quantization is usually applied to bidirectional predictive coded (B) pictures compared to intra (I) or predictive (P) coded pictures. When a difference measure is taken between a B and a P field, these quantization differences appear as image differences. This obscures the image differences of the original input signal.

Interlacing “noise” as a result of the video lines in one of the fields being vertically offset by one frame line compared to the other field. They contain different image content, even though they belong to the same original film frame. Additionally, vertical alias in the input fields hampers the distinction between actual motion and still (*i.e.*, non-moving) areas with a large amount of vertical detail. Therefore, such still areas are easily detected as areas with motion.

Hybrid content: Two or more different sources of video content may have been mixed before the signal is received by the TV set. An example is given in Fig. 2 where a 60-Hz originating ticker tape runs on top of a 24-Hz origi-



FIGURE 2 — Hybrid content consisting of 24-Hz film (converted to 60 Hz using 3:2 pull-down) and 60-Hz video (tickertape).

nating movie with 3:2 pull-down. Other examples are news broadcasts or documentaries with a presenter in video mode combined with movie material in the background, moving logo's, dynamic computer-generated content, breaking news ticker tapes, semi-transparent overlays, etc.

Special effects: For example, fade-over effects and scene cuts after pull-down introduced can break a pull-down pattern. Slow-motion scenes are usually created by means of image repetition using various pull-down patterns.

Image-comparison-based methods must balance good sensitivity for scenes with little motion with robustness to noise. Only differences above a noise threshold can be considered, but this reduces the overall detector sensitivity. This introduces the problem of finding optimal threshold settings. In Ref. 11, this is automated by computing differences for a number of thresholds, and applying simultaneous pattern matching. In Ref. 14 a statistically derived predictor depending on the current pattern and a related adaptive threshold is used to improve robustness. In general, a motion measure based on some form of motion estimation is more robust to noise, but more expensive to realize.

To handle the challenges of hybrid content and special effects, one can opt for different defensive strategies: globally processing the video's dominant mode or not applying MC-PRC. A robust solution that detects hybrid film modes and handles hybrid signals properly remains a difficult challenge. As a first step, Ref. 12 proposes a method to detect hybrid content based on a classification of different types of image-motion patterns.

4 True motion estimation

Motion estimation (ME) is the foundation of MC-PRC: the ME algorithm determines the motion trajectory of each pixel over time, generating a dense vector field. Many subsequent operations in the conversion chain depend on accurate and consistent motion vectors (MVs). As a consequence, the ME algorithm has a crucial influence on the final outcome of the conversion.

ME algorithms create an estimate of the *optical flow*, which is the *apparent* motion of the brightness pattern in the 2-D plane. The optical flow can differ from the true 2-D projection of 3-D object motion, which in some occurrences cannot be reconstructed from the image sequence due to lighting, occlusion and aperture effects.¹⁶ Note that the more accurate definition "optical flow estimation" is more frequently used in academia, yet articles in the field of picture-rate conversion mostly refer to ME or "true-motion" ME. In this section, we give a short overview of available methods and discuss their use in MC-PRC systems.

4.1 Motion-estimation methods

A ME method is defined by its match criterion, optimization/search method, and smoothness constraints. Here, we

describe three popular classes of ME algorithms, distinguished by their match criteria: (1) direct matching and correlation techniques, (2) phase plane correlation techniques, and (3) methods based on optic flow constraint equations.

Direct matching techniques are relatively straightforward and robust. The ME method determines the maximum correlation between a region (*e.g.*, a pixel, a block of pixels, or an arbitrary segment) in one image and its displaced counterpart in the associated image. All modern video coding standards (*e.g.*, MPEG-4, H.264) and many frame-rate-conversion applications are based on block-matching algorithms. The matching can be implemented in an exhaustive (full) search,¹⁷ in reduced search patterns, such as diamond search,¹⁸ or by search methods that use spatiotemporal predictions, *e.g.*, see Refs. 19–23. Various correlation criteria are used, such as the normalized cross correlation function (NCCF), the mean square error (MSE), and the sum/mean of absolute differences (SAD/MAD). The latter have the advantage that they do not require as many multiplications as the former. The matching criterion can be simplified without significantly affecting performance,^{24,25} although this is related to the search/optimization scheme used.

All direct matching techniques experience the fundamental trade-off in block size versus pronouncedness of the correlation maximum. As Fig. 3 (top row) illustrates, larger "template"-matching regions provide more robust correlation maxima compared to smaller block sizes. On the other hand, larger blocks prevent the vector field from closely following object boundaries, resulting in artifacts. Additionally, matching with larger blocks can be error prone in regions with strong deformation, zoom or rotation, as the simple translational displacement model that is inherent to block matching does not hold there. Due to these tradeoffs, block sizes of 8×8 or 16×16 pixels are found in most algorithms.

Phase plane correlation techniques exploit correlation by multiplication in the Fourier domain. Their use for dense motion estimation was pioneered in Ref. 26 and the technique has found its way into recent PRC products.²⁷ Unlike direct-matching techniques, where a block is correlated with multiple shifted versions to determine multiple correlation values, phase plane correlation (PPC) can resolve a "correlation surface" directly from two temporally non-displaced input blocks. Consequently, the block size must be (significantly) larger than the displacement (*e.g.*, block sizes of 64×64 were reported for standard-definition content). Like a full search, all motion vectors are evaluated, but at a greatly reduced cost, which is now predominantly determined by the FFT and IFFT. Correlation maxima are very sharply defined in the correlation surface due to the inherent frequency normalization. To establish dense vector fields, the technique is typically combined (hierarchically) with direct block matching on finer block grid resolutions to test and assign the MVs found in the PPC.

Optic flow constraint (OFC) equations are acquired by linearization, using the Taylor-series expansion, of the



FIGURE 3 — (Top row) Motion-vector color overlays of a FS block-matching algorithm for varying block sizes on a 720×576 input image with quarter-pixel search steps: (a) 32×32 px, (b) 16×16 px, (c) 8×8 px, (d) 4×4 px. For large block sizes, the estimation is inaccurate at object boundaries and areas with non-translational motion. For small block sizes, the correlation becomes unreliable. (Middle row) Motion vector overlays of alternative algorithms on 16×16 -px blocks: (e) phase plane correlation (PPC),^c (f) diamond search (DS), (g) predictive motion-vector-field adaptive search technique (PMVFAST),^d (h) recursive search (RS).^e (Bottom row) Motion-vector overlays of the recursive search algorithm with isotropic smoothness constraints, for varying block sizes, (i) 32×32 px, (j) 16×16 px, (k) 8×8 px, (l) motion-vector overlay of a pixel accurate optic flow method³¹ with anisotropic smoothness constraints (image and flow-driven regularization).

brightness (gradient) constancy assumption(s). The motion vectors per pixel can be calculated directly from a set of equations that contain local spatial and temporal image derivatives and global/local smoothness constraints.^{28,29} The motion vectors can be computed without a matching search, but the linearization is valid only for small motions and the derivatives per pixel can be inaccurate. Hence, strong edge-preserving regularization is needed, as well as course-to-fine-warping schemes to handle larger motions, *e.g.*, see Ref. 30. Modern OFC methods are highly accurate, but the numerical methods required to solve the large systems of equations are computationally expensive. We mention this category of methods for completeness, but we are currently not aware of commercially available optical-flow-based PRC systems. The GPU accelerated method from Werlberger *et al.*³¹ was used to compute the vector field in Fig. 3(l).

^cPPC is used to generate 10 motion-vector candidates per 64×64 -px block, which are evaluated on and assigned to 16×16 -px (sub)blocks.

^dPMVFAST algorithm with early termination disabled.

^e3DRS with 2 passes, a Gaussian random update candidate set and no penalty mechanism.

4.2 Spatial consistency

Next to the matching criteria, the most important aspect of PRC-oriented ME algorithms is their approach to spatial vector field consistency. As Fig. 4 illustrates, algorithms that are based solely on residue minimization are not suitable for PRC. For high-quality conversion, the “true” motion is needed, which is spatially consistent within objects. To gen-



FIGURE 4 — Interpolation results of (left) full search and (right) “true” motion estimation using recursive search. Algorithms-based-only residue optimization are generally unsuitable for PRC, as they generate vector fields that are optimal in terms of match error, yet contain many errors in areas with repetitive detail.

erate piece-wise smooth vector fields, we distinguish three approaches:

Methods based on (limited) spatio-temporal candidate sets. These are typically employed in direct-matching algorithms. By limiting the set of motion-vector “candidates” tested to those (close to) neighboring blocks, the computational efficiency, and the consistency of the motion vector fields is improved, *e.g.*, see Refs. 19 and 20.

Methods based on post-filtering. In separate filtering passes, independent from the motion estimation, outliers are removed from the vector field, *e.g.*, see Ref. 32. Post-processing methods are typically non-linear (*e.g.*, median, or based on neighborhood correlation quality observations) and are often employed to improve the consistency of an exhaustive search method. Alternatively, they are used to improve vector fields acquired from compressed video streams, see, *e.g.*, Ref. 33.

Methods based on additional energy terms. An energy term that describes the smoothness, *i.e.*, the difference of the tested MV to the MVs of neighboring blocks, is added to the matching criterion. This approach is predominantly used in optic flow methods, and requires an iterative optimization. Variants in direct-matching are possible.²³

The bottom row of Fig. 3 illustrates the effect of smoothness constraints on the output vector fields (illustrated are the recursive search method for varying block sizes, and the optic flow method from Ref. 31). The dependency on larger block sizes for a higher correlation quality is reduced.

4.3 Implementation

High-quality ME is a computationally demanding task and challenging to implement on low-cost, low-power embedded platforms. Particularly for real-time block-matching methods, much effort has been spent on techniques that reduce the memory bandwidth and operations count. This is apparent in the reduced search schemes that evaluate (*i.e.*, determine the correlation value of) only a small number of motion-vector “candidates” per block instead of all possible motion vectors in exhaustive searches.^f By selecting this set of “candidate” MVs based on the MVs of spatiotemporally neighboring blocks, a relatively small number of correlation matches (*e.g.*, five or less^g) per block are adequate to determine the vector fields with sufficient quality for PRC. In popular video codecs, methods based on “predictors” are used that similarly exploit the principle of a search started around the MVs of neighboring blocks, *e.g.*, see Ref. 20.

Although, efficient search schemes have drastically reduced computational and memory bandwidth demands, from a memory perspective there is one remaining issue:

^fA full-search ME with quarter pixel accuracy and a 100-px square search window computes 1.6e5 block correlations per block.

^gReference 21 describes a recursive search implementation that uses on average less than 2.7 block correlations per block.



FIGURE 5 — Interpolation results of a vertically moving sequence with (left) small and (right) large velocity range.

large velocity ranges. Supporting a larger velocity range does not pose significant algorithmic challenges, but it does increase the architectural challenges as local buffer sizes have to be increased. In most architectures, this especially impacts the vertical velocity range, as the video stream enters the system in a line-based fashion, and hence is also buffered that way. The consequence of an insufficiently large velocity range is shown in Fig. 5, where break-up artifacts become visible (right).

A compromise in system cost and picture quality may be achieved by incorporating detection mechanisms when the velocity range is exceeded, and reverting to global fallback in such cases, as detailed in Sec. 7. Although this is not an ideal solution, it may be preferred over the visually disturbing artifacts that occur if the velocity range is exceeded.

5 Robust interpolation

Given “true” motion vectors, the subsequent interpolation may seem to be the case of averaging motion-compensated pixels. However, we cannot assume that motion vectors are always correct, as otherwise highly visible artifacts may occur, illustrated in Fig. 6, left. This means that additional robustness measures are required in the interpolation. These are typically based on the mismatch between both motion-compensated pixels, and the mismatch between both non-motion-compensated pixels.

The underlying assumption is that if the mismatch of the motion-compensated pixels is high, either in absolute terms or relative to the non-motion compensated pixels, the motion vector may have been unreliable. In such cases, a local fallback is used, which is often the non-motion compensated average.

The rationale behind this is twofold. Firstly, the non-motion-compensated average results in a temporal averaging of the images, which is a perceptually unobtrusive artifact. Secondly, interpolation errors are most visible at the borders of non-moving areas (such as subtitles or logos), and the above choice results in a natural bias towards the correct interpolation of such areas. Figure 6 illustrates this improvement with the example of the so-called cascaded median.³⁹

In general, the cascaded median belongs to the group of rank-order-based filters, which are well suited for robust



FIGURE 6 — Interpolation results using (left) motion-compensated average and (right) cascaded median.

interpolation. A rank-order filter selects a pixel of a specified rank within an ordered set by the intensity of the pixels in the filter's aperture. By specifying a rank towards the middle of the range, *e.g.*, 0.5, the 50th percentile, outliers are naturally rejected in the filter output and the output of the filter is the pixel with the median value.

In this respect, it should be noted that the aperture of the filter is primarily temporal, *i.e.*, containing motion-compensated and non-motion-compensated pixels from the previous and the current frame. However, to account for slight inaccuracies of the motion vector, the aperture may extend in the spatial dimension, *e.g.*, containing the pixels in a window around the motion-compensated pixels.

Therefore, many variations of the cascaded median are possible. It should be noted, however, that specifying a single rank results in so-called switching artifacts, as slight variations in pixels within the filter aperture result in a different pixel being selected as the output pixel. To prevent such switching artifacts, a range of ranks can be selected such that all pixels having a rank within the range are averaged, or weighted with a non-zero weight. This provides a smoothing of the output of the filter, yet by its nature increases the contribution of outliers to the filter output. Hence, a trade-off has to be determined that sufficiently rejects outliers and smoothes switching artifacts.

Furthermore, in the field of video coding, overlapped-block motion-compensation (OBMC) is frequently used to avoid highly visible blocking artifacts due to a limited resolution of the motion vector field.⁴⁰ Here, the motion vectors of neighboring blocks are used in the interpolation of the current block. The multiple sets of motion-compensated pixels are then weighted, *e.g.*, based on an intensity rank or the spatial distance of the output pixel to the block for which the motion vector was determined.

As such, OBMC forms an alternative to estimating a high-resolution motion-vector field by using small blocks and is similar in result to performing motion estimation with overlapped blocks. It also is an alternative to otherwise post-processing the motion-vector field to prevent blocking artifacts in the interpolation, *e.g.*, see Ref. 8.



FIGURE 7 — Interpolation with and without occlusion handling: (left) cascaded median interpolation, (right) explicit occlusion detection and correction.

6 Occlusion handling

Special care has to be taken in the handling of borders of moving objects. Here, an effect called occlusion is occurring, in which image parts are only visible in either of the two adjacent video frames. Although this effect is well known in the literature, its solution is not trivial in practice. Earlier products incorporating MC-PRC did not apply special measures in occlusion areas, other than having a robust interpolation, *e.g.*, by means of the cascaded median,³⁹ detailed in the previous section. However, if the background object contains detail, a shadow-like artifact in the occlusion areas, when observed in real-time, is perceived as a halo around moving objects. This effect is shown in Fig. 7.

To fundamentally solve the occlusion/halo artifacts, the algorithm has to perform a number of steps: (1) the occlusion areas have to be located, (2) the correct motion vectors have to be propagated into these areas, and (3) the interpolation has to be adopted to use data from only one of both video frames. The result of such an algorithm³⁴ is shown in Fig. 7, right, which illustrates the clear reduction of occlusion/halo artifacts.

The occlusion detection is a challenging task as the detection must guard for overclassification in the foreground object, while preventing underclassification of the real occlusion areas. The correction of falsely detected occlusion in the foreground object can yield highly visible errors, *e.g.*, Fig. 8, whereas overclassification in the background area typically does not result in visible artifacts.



FIGURE 8 — (Left) Incorrect occlusion handling can yield visible artifacts; most notable is "foreground eat-a-way" where parts of the foreground object are destroyed. (Right) the subsequent non-interpolated frame.

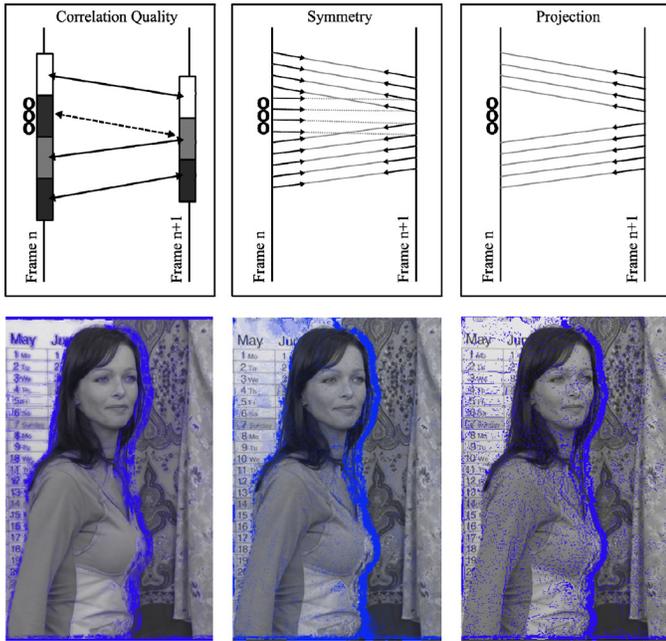


FIGURE 9 — Occlusion information is typically derived from (left) the matching error, (middle) the MC difference between the forward and backward motion vector fields, or (right) the projection of a single motion vector field. Note that the first two are continuous measures, whereas the latter is discrete (binary).

We distinguish three common detection methods based on: image matching, vector-field projections, and vector-field forward-backward comparisons, illustrated in Fig. 9. A more extensive overview, albeit for a stereo-images case, is presented in Ref. 35. A benchmark of some methods in the context of picture-rate conversion can be found in Ref. 38.

Image-matching-based methods (Fig. 9, left) are based on the motion-compensated comparison of pixels or image patches, *e.g.*, by means of thresholding the correlation values of the final motion vectors after (or during) the motion estimation. The rationale is that image patches that do not have a good match between two consecutive pictures must be the result of occlusion. This assumption does not always hold, *i.e.*, most ME algorithms will fundamentally try to minimize the correlation value in occlusion areas by using “incorrect” vectors. This makes it difficult to distinguish occlusion areas based only on correlation quality.

Forward-backward comparison-based methods (Fig. 9, middle) are based on the motion-compensated comparison of motion vectors from two vector fields, *i.e.*, a forward motion estimate from frame n to $n + 1$ and a backward motion estimate from frame $n + 1$ to n . Ideally, both should match in non-occluded regions. However, the two motion estimates can deviate in areas where the motion is ambiguous due to the aperture problem or lack of texture, causing incorrect occlusion detection. A “coupled” motion estimation of the two vector fields, *e.g.*, Ref. 37 can address this but is more expensive in implementation compared to non-coupled estimates.

Projection based methods (Fig. 9, right) use holes or double assignments in a motion-compensated vector field as an indication for occlusion. For example, if a motion vector field is projected forward in time by one frame, certain pixels will not be “hit” by a motion vector. These pixels indicate uncovered areas. In Ref. 36 this approach is refined using density measurement.

A combination of the above detection principles is possible, *e.g.*, see Ref. 38.

Lastly, we note that the quality of the occlusion detection depends foremost on the quality of the motion estimates. If the motion estimates are “perfect”, simple occlusion detection mechanisms suffice. Unfortunately, this is not often the case.

7 Global fallback

The previously described components can handle the majority of video content, but situations can still occur in which MC-PRC generates highly visible artifacts. This mainly holds for scenes that contain extremely irregular or erratic motion, or in which consecutive video frames show very little correspondences. The latter may be the case due to film edits (*e.g.*, scene changes) or in synthetically generated content.

For such video content, picture-rate conversion using simple picture repetition is typically preferred over MC-PRC, as interpolation artifacts disappear, only to be replaced by judder. Perceptually, the human observer will have difficulty tracking the motion, making the re-appearance of judder a relatively minor issue, whilst interpolation artifacts tend to remain visible in such cases. The effects of this are illustrated in Fig. 10.

In the decision when to temporarily revert to picture repetition, multiple statistical cues may be used. For example, the noisiness of the motion-vector field and/or the average residue can be used, as both may indicate the presence of large motion-vector errors.

An example of the statistical cue of motion-vector field noisiness is given in Fig. 11, where the noisiness is plotted as a function of the frame number. This particular sequence includes a video segment A containing a very fast panning motion, video segments B and C containing rapid and erratic



FIGURE 10 — Interpolation results in the case of erratic motion without (left) and with (right) global fallback.

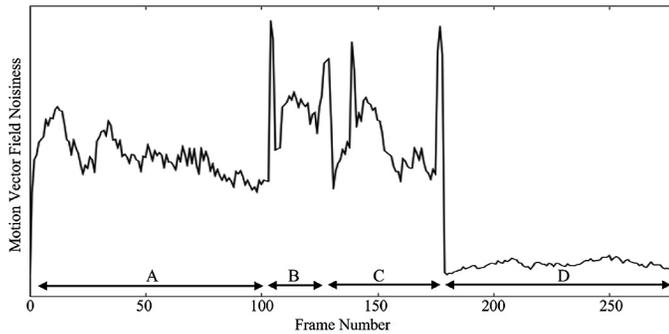


FIGURE 11 — Noisiness of motion vector field as a function of frame number for four consecutive video segments A, B, C, and D.

motion, and a video segment D segment containing a slow panning motion.

Clearly, these segments and the corresponding changes between the segments, also known as scene changes, can be recognized from the noisiness measurement. In particular, the scene-changes appear as clearly defined spikes in the noisiness measure, making the identification of such scene changes and therefore the triggering of global fallback relatively effortless.

Similarly, it can be seen that, for example, segment B has an average motion-vector noisiness that is higher than that of segment A. In fact, the interpolated results in this segment are likely to contain visible artifacts. It may therefore be a good design choice to trigger fallback for these types of segments next to the fall-back during scene changes.

As the exact tuning of global fallback involves a trade-off between different types of artifacts, *i.e.*, in this case of judder versus interpolation artifacts, it is typically as much a matter of subjective preference as it can be objectively determined from statistical cues.

Lastly, to prevent visible-switching artifacts, the transition in and out of fallback may be performed gradually.⁴¹

8 Conclusions

MC-PRC plays a key role in the improvement of the motion portrayal of flat-panel displays. However, the label “MC-PRC” does not automatically guarantee a better picture quality, as a wide variety of requirements has to be met in order to handle a broad range of video content. This includes content with, for example, occlusions, ambiguous motion, very large velocities, and different pull-down patterns. Correct handling of this content is of paramount importance as artifacts of incorrect MC interpolation are typically perceived as more annoying than motion judder as a result of applying no MC interpolation.

Consequently, current medium-to-high-end picture-rate-conversion systems contain a highly “tuned” set of components which provide, next to high-quality motion estimation, support for film mode detection, occlusion robust interpolation, and a “soft” global fallback. Robustness measures and fallback scenarios are vital to prevent

visible interpolation artifacts, as every component in the processing chain has the potential to fail on certain types of content.

Additionally, the design poses significant architectural challenges, as ultimately the PRC must be implemented on a low-cost, low-power platform. For example, the application of cost savings in the form of smaller buffer sizes (which impact the support for large velocities) and simplified designs for film mode detection (which impact the support for different content) is a non-trivial process.

Acknowledgment

The authors would like to thank Thijs Withaar from NXP Semiconductors for his contributions to this work.

References

- 1 D. Parker, “The dynamic performance of CRT and LC displays,” *Display Systems Design and Applications*, 353–364 (1997).
- 2 H. Okumura, “A new low-image-lag drive method for large size LCTVs,” *J. Soc. Info. Display* **1**, No. 3, 335–339 (1993).
- 3 G. de Haan and R. Braspenning, “Video scanning format conversion and motion estimation,” in *The Digital Signal Processing Handbook*, 2nd edn., V. K. Madiseti (ed.) (CRC Press LLC, 2009), pp. 16/1–16/43.
- 4 J. Watkinson, “The Engineer’s Guide to Motion Compensation,” *Handbook Series* (Snell & Wilcox, 1994).
- 5 G. de Haan, *et al.*, “IC for motion compensated 100-Hz TV, with a smooth motion movie-mode,” *ICCE Digest of Technical Papers*, 40–41 (1995).
- 6 “Global DTV platform TV550,” retrieved on February 10th 2010 from http://www.nxp.com/applications/tv/digital_tv/tv550/.
- 7 S. H. Han *et al.*, “Converting the interlaced 3:2 pull-down film to the NTSC video without motion artifacts,” *Proc. ICIP* (2005).
- 8 R. J. Schutten and G. de Haan, “Real-time 2-3 pull-down elimination applying motion estimation/compensation in a programmable device,” *IEEE Trans. Consumer Electron.* **44**, No. 3, 930–938 (1998).
- 9 C. C. Ku and R. K. Liang, “Robust layered film-mode source 3:2 pull-down detection/correction,” *IEEE Trans. Consumer Electron.* **50**, No. 4, 1190–1193 (2004).
- 10 P. D. Swartz, “Film source video detection,” Faroudja, Patent No. WO 99/20040, Oct. 10, 1997.
- 11 A. K. Riemens and A. Domnisse, “Motion sequence pattern detection,” Philips Electronics, Patent No. WO 2004/049698 A1, Nov. 26, 2002.
- 12 A. K. Riemens *et al.*, “Recognizing film and video occurring in parallel in television fields,” Philips Electronics, Patent No. WO 2004/054256 A1, Dec. 12, 2002.
- 13 M. C. Capps, “Method for detecting film pulldown cadences,” Texas Instruments, Patent No. US 2008/158414 A1, Dec. 29, 2006.
- 14 D. Znamenskiy and C. N. Cordes, “Film cadence detection,” Philips Electronics, Patent No. WO 2008/081386 A1, Jan. 3, 2007.
- 15 M. C. Liu *et al.*, “Real-time MPEG video encoder with embedded scene change detection and telecine inverse,” *ICCE Digest Technical Papers*, 246–247 (2002).
- 16 A. Tekalp, *Digital Video Processing* (Prentice Hall Signal Processing Series, 1995).
- 17 J. R. Jain and A. K. Jain, “Displacement measurement and its application in interframe coding,” *IEEE Trans. Commun.* **29**, No. 12, 1799–1808 (1981).
- 18 S. Zhu and K. K. Ma, “A new diamond search algorithm for fast block matching motion estimation,” *IEEE Trans. Image Processing* **9**, 287–290 (2000).
- 19 G. de Haan *et al.*, “True-motion estimation with 3-D recursive search block matching,” *IEEE Trans. Circuits Systems Video Technol.* **3**, No. 5, 368–379 (1993).
- 20 A. Tourapis *et al.*, “Highly efficient predictive zonal algorithms for fast block-matching motion estimation,” *IEEE Trans. Circuits Systems Video Technol.* **12**, No. 10, 934–947 (2002).

- 21 N. Atzpadin *et al.*, "Stereo analysis by hybrid recursive matching for real-time immersive video conferencing," *IEEE Trans. Circuits Systems Video Technol.* **14**, 321–334 (2004).
- 22 G. Lee *et al.*, "Algorithm/architecture co-design of 3-D spatio-temporal motion estimation for video coding," *IEEE Trans. Multimedia* **9**, No. 3, 455–465 (2007).
- 23 C. Bartels *et al.*, "Smoothness constraints in recursive search motion estimation for picture rate conversion," To appear in *IEEE Trans. Video Technology* (accepted).
- 24 R. Srinivasan and K. R. Rao, "Predictive coding based on efficient motion estimation," *Proc. ICC*, 521–526 (1984).
- 25 C. N. Cordes and E. B. Bellers, "Efficient true motion estimation using integral frame attributes," *ICCE Digest Technical Papers*, 1–2 (2009).
- 26 G. A. Thomas, "Television motion measurement for DATV and other applications," *BBC Research Report*, No. BBC RD 1987/11 (1987).
- 27 AMD Xilleon Panel Processors Press Presentation (2008), retrieved on February 10th 2010 from <http://www.amd.com>.
- 28 B. Horn and B. Schunck, "Determining optical flow," *Artificial Intelligence* **17**, No. 1–3, 185–203 (1981).
- 29 B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," *Proc. Imaging Understanding Workshop*, 121–130 (1981).
- 30 T. Brox *et al.*, "High accuracy optical flow estimation based on a theory for warping," *Proc. ECCV* (Springer-Verlag, Berlin, 2004).
- 31 M. Werlberger *et al.*, "Anisotropic Huber-L1 optical flow," *Proc. British Machine Vision Conf.* (2009).
- 32 B. Choi *et al.*, "New frame rate up-conversion using bi-directional motion estimation," *IEEE Trans. Consumer Electron.* **46**, 603–609 (2000).
- 33 G. Dane, "Temporal frame interpolation by motion analysis and processing," Ph.D. Thesis University of California (2005).
- 34 E. B. Bellers *et al.*, "Solving occlusion in frame-rate up-conversion," *IEEE Digest Tech. Papers ICCE*, 1–2 (2007).
- 35 G. Egnal and R. Wildes, "Detecting binocular half-occlusions: Empirical comparisons of five approaches," *IEEE Trans. Pattern Analysis Machine Intelligence* **24**, No. 8, 1127–1133 (2002).
- 36 S. Ince and J. Konrad, "Geometry-based estimation of occlusions from video frame pairs," *Proc. ICASSP*, 933–936 (2005).
- 37 C. Bartels and G. de Haan, "Temporal symmetry constraints in block matching," *Proc. ISCE*, 749–752 (2009).
- 38 C. Bartels and G. de Haan, "Occlusion classifiers for picture rate conversion," *IS&T/SPIE Electronic Imaging*, Paper 7257-49 (2009).
- 39 O. A. Ojo and G. de Haan, "Robust motion-compensated video up-conversion," *IEEE Trans. Consumer Electron.* **43**, No. 4, 1045–1056 (1997).
- 40 B.-D. Choi *et al.*, "Motion-compensated frame interpolation using bilateral motion estimation and adaptive overlapped block motion," *IEEE Trans. Circuits Systems Video Technol.* **17**, No. 4, 407–416 (2007).
- 41 O. A. Ojo and H. Schoemaker, "Adaptive global concealment of video up-conversion artefacts," *IEEE Trans. Consumer Electron.* **47**, No. 1, 206–207 (2001).



Chris Bartels received the M.Sc. degree in electrical engineering from Eindhoven University of Technology in 2006. He is currently pursuing his Ph.D. degree at the same institute on a grant from Philips. His research focus is on algorithms for real-time video processing; in particular, motion estimation and frame-rate-conversion techniques.



Nico Cordes received his M.Sc. degree in electrical engineering from the Eindhoven University of Technology in 2003. After graduating, he started his career at Philips Research, where he designed video-enhancement algorithms for OLED displays. In 2005, he switched to the area of video-format conversion, where his research was focused on frame-rate-conversion algorithms. In 2008, he continued his activities at NXP Semiconductors. His work has resulted in 13 published patent applications.



Bram Riemens received his B.Eng. degree in electrical engineering from Fontys Hogescholen in Eindhoven. He joined Philips Research in 1984 and moved to NXP Semiconductors in 2006. Since 1994, his research focuses on algorithm-architecture co-design of advanced video-enhancement functions in real-time embedded systems-on-chip. He contributed to commercially available ICs, published several papers, and holds about 35 international patents and patent applications.



Gerard de Haan received his M.Sc. and Ph.D. degrees from Delft University of Technology. He joined Philips Research in 1979, where he is currently a Research Fellow active in the area of video processing and analysis. Since 2000, he also holds the chair "Video processing for multimedia systems" at Eindhoven University of Technology. His research resulted in various commercially available ICs, three books, two handbook chapters, 145 papers, and more than 100 international patents and patent applications. He received five best paper awards, the IEEE Chester Hall Award, the Gilles Holst Award, bronze, silver and gold patent medals, while his work on motion received the EISA European Video Innovation Award, and the Wall Street Journal Business Innovation Award.