Robust motion estimation design methodology

Citation for published version (APA):

Heinrich, A., Bartels, C. L. L., Vleuten, van der, R. J., & Haan, de, G. (2010). Robust motion estimation design methodology. In *Proceedings of the 2010 Conference on Visual Media Production (CVMP), 17-18 November 2010* (pp. 49-57). Institute of Electrical and Electronics Engineers. <https://doi.org/10.1109/CVMP.2010.14>

DOI:

[10.1109/CVMP.2010.14](https://doi.org/10.1109/CVMP.2010.14)

Document status and date:

Published: 01/01/2010

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

ROBUST MOTION ESTIMATION DESIGN METHODOLOGY

A. Heinrich¹, C. Bartels², R.J. van der Vleuten¹, G. de Haan¹

¹ Philips Research Laboratories, Eindhoven, The Netherlands, adrienne.heinrich@philips.com

² Eindhoven University of Technology, The Netherlands

Abstract

For motion-adaptive video retiming methods, the quest to lower the implementation complexity and improve the quality of motion estimation algorithms still continues. Comparing different motion estimators (MEs) and/or fine-tuning ME parameters is a time-consuming task, and it is even more demanding to identify the MEs with a robust performance among all the well-performing MEs. Therefore, a computer-aided design methodology is required to effectively explore the large design space. Such a methodology requires objective performance metrics. As it is hard to find perfect metrics, we present a design methodology that can use suboptimal measures and still identify robust MEs. The proposed methodology is demonstrated using two different MEs.

Keywords: Motion estimation, Retiming, Design methodology, Performance measure, Metrics

1 Introduction

Retiming video sequences requires vectors describing the local motion between consecutive images in a video sequence. These vectors are typically obtained by performing motion estimation. Improving the quality of the motion estimation algorithm and lowering the implementation complexity are important factors in the design of a motion estimator (ME).

Comparing different motion estimators (MEs) with each other and/or fine-tuning their parameters is often a time-consuming and complex task. Even more demanding, and yet one of the most important objectives in practice, is to identify robust MEs in terms of a good average performance with a rather small performance variation for different challenges. For the application of retiming, metrics have been developed to assess the performance of a ME, e.g. [1, 2]. However, despite all the developed performance measures, the metrics still do not perfectly correspond to the quality that is perceived by human observers. Thus, it is required that the ME design methodology can deal with suboptimal metrics and still identify robust MEs in an automated manner.

In this paper, the proposed methodology has been successfully applied to two different ME types: Recursive Search block matching (RS) and Phase Plane Correlation (PPC). Both are relevant motion estimators for the application of retiming and are commercially available in products (others could have been chosen as well). Spatio-temporal prediction methods such as RS, e.g. [3, 1, 4], are applied in practice (e.g. [5, 6]), and so are alternatives based on PPC [7, 8].

In Section 2, we introduce the proposed design methodology where the test data selection, the performance measures and the identification of robust ME settings are addressed. The design methodology is applied in Section 3 to both a RS approach and PPC. Section 4 discusses the validation of the proposed methodology regarding suboptimal performance measures and robustness, and provides a benchmark with other techniques. Section 5 summarizes our conclusions.

2 Design Methodology

In order to automatically identify parameter settings of robust MEs for retiming video sequences, we present a methodology that can successfully deal with performance measures that are suboptimal in the sense that they do not fully reflect the perceived video quality. A three-step approach is suggested where, firstly, the variety of conditions under which the MEs should perform well is defined and appropriate test data is selected. Secondly, a contour line or trade-off curve illustrates the achieved compromise between the motion vector prediction accuracy and consistency. Thirdly, an attractive segment is identified containing all MEs within a defined distance from an attractive section of the contour line. Histogram analysis provides the distribution of MEs within the attractive segment to identify the parameter settings of the MEs that are least sensitive to varying settings and thus most robust.

The variety of performance conditions is addressed in Section 2.1. Section 2.2 discusses the trade-off curve of useful metrics for the application of retiming and the methodology's effectiveness in dealing with potentially suboptimal measures. Among the well-performing MEs, robust MEs and their parameter settings are identified in

Figure 1: Test sequences used for the quantitative evaluation. Note that the bottom row is reused for still images and de-interlacing with a typical de-interlacer, e.g. [9].

Section 2.3.

2.1 Test Data Selection

A ME should perform well under all considered conditions for the retiming application. These conditions are included in the test data which should address ME challenges such as repetitive structures, small objects, subtitles and ticker tapes, several layers with different motion, de-interlaced images with typical de-interlacers of average quality (e.g. [9]), large motion, and occlusion areas. To ensure a satisfactory performance with less challenging test material, also fairly straightforward sequences for ME should be included, as well as a repeated still image. The test data were selected according to these criteria. A snapshot of each Full-HD test sequence is shown in Fig. 1.

We expect a well-performing ME to have a good average performance for all challenges. For individual challenges, we acknowledge that other ME parameter settings may render a better result, however, the objective in the ME design for retiming video sequences remains a good overall performance. Therefore, the average performance over all test sequences is compared.

2.2 Performance Measures

Two fundamental characteristics are recognized as the basis of ME design: The brightness constancy assumption when the true motion is found and the smoothness constraints to enforce consistent motion fields within a moving object. The trade-off between smoothness terms and brightness constancy in the form of luminance comparisons has already become apparent in the early optical flow implementations [10]. It is also recognized in [1] and [2], that accurate predictions at a highest possible consistency are necessary for a satisfactory viewing experience. Relevant metrics addressing the prediction accuracy, temporal continuity and spatial

consistency of the motion vectors are documented in [1] and [2]. The prediction accuracy and temporal continuity are quantitatively assessed with the ‘M2SE’ [2],

$$\text{M2SE}(n) = \frac{1}{n_h \cdot n_w} \sum_{\vec{x} \in W} (F_o(\vec{x}, n) - F_i(\vec{x}, n))^2, \quad (1)$$

and the spatial inconsistency measure ‘SI’ based on [2],

$$\text{SI}(n) = \sum_{\vec{x}_b \in W_b} \sum_{k=-1}^1 \left(\frac{|\Delta_x(\vec{x}_b, k, l, n)| + |\Delta_y(\vec{x}_b, k, l, n)|}{8 * N_b} \right), \quad (2)$$

where n_h and n_w are the image height and width in pixels, respectively, W is the set of all the pixels in the entire image, $F_o(\vec{x}, n)$ the luminance of the original image at position $\vec{x}$ and at the temporal position n . F_i is the motion compensated average of frames $n-1$ and $n+1$ by applying the vectors estimated for frame n , $\vec{x}_b$ the position of the block b among the set of all the blocks W_b in the entire image, N_b the number of blocks in an image and

$$\Delta_x(\vec{x}_b, k, l, n) = d_x(\vec{x}_b, n) - d_x(\vec{x}_b + \begin{pmatrix} k \\ l \end{pmatrix}, n), \quad (3)$$

$$\Delta_y(\vec{x}_b, k, l, n) = d_y(\vec{x}_b, n) - d_y(\vec{x}_b + \begin{pmatrix} k \\ l \end{pmatrix}, n), \quad (4)$$

where d_x and d_y are the computed motion vectors. Different block sizes in the SI metric, e.g. 8x8-pixel blocks vs. 1x1-pixel blocks, return different results due to the metric bias towards MEs with larger blocks than employed in the metric, thus appropriate block dimensions should be chosen for the set of MEs one would like to compare.

The M2SE measures how well the interpolation result corresponds to true motion using temporally extrapolated motion vectors, whereas the SI indicates the spatial smoothness of the computed motion field. The motion field and interpolated images are evaluated after performing ME on the second image of the input sequence since the pixels from a previous image are included in the M2SE computation. The corresponding PSNR is calculated from the M2SE: $\text{PSNR}(n) = 10 \cdot \log_{10}((2^{\text{NB}} - 1)^2 / \text{M2SE}(n))$, where NB is the number of bits used for representing the video data.

Since both a high PSNR as well as a consistent motion field are characteristics of a good ME, the *PSNR - Consistency* plot as shown in Fig. 2 is introduced as a means to capture the achieved PSNR performance in relation to the consistency of the motion field. The inverse mean of the PSNR and the mean inconsistency values (SI) are plotted by computing the average performance of all parameter setting combinations with regard to the different test sequences. The optimal ME with a high PSNR and a low inconsistency is located in the bottom left corner. We call a ME which is not surpassed by any other ME in both regards (consistency and PSNR) an ‘optimal’ ME. This

set of optimal MEs lies on the ‘contour line’ or ‘trade-off curve’ as described in [11] and the range of best MEs is found close to this line. The blue contour line in Fig. 2 shows the compromise between PSNR and consistency performance. The area of possible MEs is indicated by the green shading in Fig. 2.

Figure 2: The PSNR-Consistency trade-off graph of the ME design space where the green shaded area indicates the area of possible MEs which are bounded by the contour line (blue). The black dashed lines indicate the minimal PSNR and maximal Inconsistency for the attractive contour line section. The range of optimal MEs in the attractive segment are highlighted by the red arrows.

Slightly different metrics than the ones suggested here are developed and employed in other research work, e.g. [1]. It is apparent that these metrics share the same emphasis on prediction accuracy, temporal continuity and spatial motion field consistency. However, these features can be represented in different ways, particularly since the quest for the optimal metric is yet unresolved. Due to the close similarity of the considered metrics, we expect that the MEs on the contour lines obtained with different metrics may be different, but that their deviation from the location in one PSNR-Consistency plot compared to another PSNR-Consistency plot is minor.

2.3 Identification Of Robust Motion Estimator Settings

We have observed that not all MEs on the trade-off curve lead to visually pleasing retimed video sequences, e.g., zero motion field inconsistency at the left of the curve can be achieved with zero vectors, which basically reduce retiming to frame repetition resulting in perceived judder. Therefore, retimed sequences with MEs lying on the contour line are compared and the following related retiming benefits are observed from low to high PSNR.

- Better alignment of motion vectors with the edge of moving objects, resulting in reduced artifacts in occlusion regions
- Improved temporal consistency of the motion field, resulting in reduced flickering
- Higher convergence speed (for the steep part of the contour line)

Along the contour line, an improvement of the motion vectors regarding the object alignment (clearly better with an ME lying on the intersection of the horizontally dashed line and the contour line in Fig. 2) is visible which causes less occlusion artifacts than with MEs with a smaller PSNR. A turning point is perceived where the performance degrades with larger motion field inconsistencies, starting on the left of the vertically dashed line in Fig. 2 and continuing beyond the dashed line. The MEs in that area are perceived as worse due to slightly disturbing flickering artifacts. Beyond the vertically dashed line, the MEs become unacceptable. The relevance of the spatial inconsistency metric is confirmed since a high PSNR at the cost of consistency is not preferred.

Therefore, further analysis is limited to an attractive segment of MEs in the PSNR-Consistency trade-off plot. Firstly, an attractive contour line section is defined and secondly, in order to allow for potentially suboptimal metrics, a distance from the attractive contour line section is determined within which all the MEs are treated as equally optimal.

The interesting section of the contour line is limited to MEs with a good visual quality. However, due to metric inaccuracies, not a too narrow section of the contour line is chosen. An attractive section of the contour line is suggested where the SI limit is set to 8 as is indicated by the vertically dashed line in Fig. 2 and where a PSNR of at least 27.8 is required (see horizontally dashed line in Fig. 2). A large improvement in visual quality is perceived with MEs within the attractive section compared to MEs outside the attractive section and among the MEs within the attractive section only slight performance differences are apparent. There the PSNR hardly increases and room for a rather large improvement in terms of consistency is given.

In order to allow for potentially suboptimal performance measures as discussed in Section 2.2, also the MEs within a particular distance from the contour lines should be investigated. The range of optimal MEs in the attractive segment is thus extended as shown with the red arrows and red dashed line in Fig. 2. Among the MEs with a satisfactory performance, the distribution of the parameter settings is analyzed and their values compared in a histogram analysis where the (normalized) count for each parameter setting is saved. The most robust MEs are identified by high counts since a ME parameter setting

that yields more often an optimal ME is preferred as it is assumed less sensitive to varying settings of other parameters. Therefore, the histogram of the parameter settings of a particular parameter is analyzed and among the high counts the one with the setting closest to their expectation value is selected.

3 Case Study On Motion Estimators

The methodology developed in Section 2 is applied to two different types of MEs. A block-based RS motion estimation method is used in Section 3.1 where a search is conducted for the most probable motion vectors obtained from the hierarchical-spatio-temporal neighborhood and random variations of few motion vector candidates. In Section 3.2 the same analysis is performed with a phase plane correlation (PPC) ME.

3.1 Recursive Search Motion Estimation

We investigate a hierarchical ME approach using resolution down-scaling, which we call multi-scale block-matching ME. Using down-scaling, the coarser motion vectors are obtained from block-matching at a lower spatial resolution and can be successively refined at higher resolutions. We will combine the multi-scale approach with a hierarchical ME method known as multi-grid block-matching [12]. In this method, a coarse motion vector is first found using a large block size and this vector is successively refined for the smaller blocks into which the larger blocks are decomposed (using a quad-tree decomposition). By combining the multi-scale and multi-grid approaches, we aim at reducing the computational complexity and are flexible in investigating the effects of using different block sizes and scale factors.

The multi-scale and multi-grid approach are illustrated by the scale pyramid shown in Fig. 3, where ME is performed on higher scales at the top of the pyramid first and motion vectors are propagated down the pyramid to the lower scales by means of hierarchical candidates.

The block-matching method we apply is the RS ME of [2]. In contrast to the usual RS candidate structure of [2], the temporal candidate is closer to the current block for all the hierarchical approaches because, for coarse scales, the temporal candidate may come to lie outside the object in which the current block is located.

The PSNR-Consistency plot and the contour line of the optimal RS MEs are given in Fig. 4 and Fig. 5. Note that the SI measure is computed based on 8x8-pixel blocks. The MEs within the attractive segment are shown in Fig. 6. Based on the parameter histogram analysis, which is elaborately discussed in [13], 62 multi-scale MEs have been identified out of the 1745 MEs in the attractive segment (see Fig. 6). An overview of the proposed parameter settings of this ME type is given in Table 1 where block settings and performance are rendered. The

first row of Table 3 shows the mean performance for the 62 robust MEs. From the expectation value of the block dimension distributions of the 62 MEs we determined the proposed ME settings given in the second row. The resulting ME happens to coincide with Opt. ME 6 in Fig. 5 which was visually perceived as the most pleasing ME among the seven MEs on the contour line.

3.2 Phase Plane Correlation Motion Estimation

PPC was developed in the '80s [7] and is employed in state-of-the-art products [8]. Instead of obtaining motion vector candidates from a spatio-temporal neighborhood as in RS, PPC retrieves the motion vectors by performing phase correlation in the Fourier domain between spatially corresponding blocks from consecutive images. A correlation plane of displacement peaks is returned of which a subset is used as motion vector candidates in a consequent block matching operation on smaller blocks. Among the most dominant peaks in the obtained displacement field, the peak yielding the minimal match error between the motion compensated and the original smaller blocks is selected.

We implemented PPC-based MEs based on [7] (which may not reflect current product implementations) with the parameter variations as given in Table 2. In total, 1800 ME parameter combinations are investigated. Initially, a two-dimensional Fourier transform is performed on the larger blocks with dimensions $m_l \times m_l$. The n_p most dominant displacement peaks are considered motion candidates for the smaller blocks with dimensions $m_s \times m_s$. Another parameter is the block step size s_b based on which the pixel locations of the next $m_l \times m_l$ block are selected for the next FFT operation. The values for s_b are set in the range between the largest m_s setting (16) and the current m_l dimension. The displacement of the larger blocks can be determined with pixel or sub-pixel accuracy. The binary variable a_s indicates a sub-pixel accuracy of 0.25 pixels when $a_s = 1$ and pixel accuracy when $a_s = 0$.

The contour line (with the 8x8 block-based SI measure) of the PPC MEs, the overall PSNR-Consistency plot, and the attractive segment of optimal MEs are given in Fig. 5, Fig. 7 and Fig. 8.

The histogram analysis is conducted for all parameters in Table 2. Among the 54 MEs within the attractive segment, we found that the block dimensions of both the larger and smaller blocks converge to $m_l = 128$ and $m_s = 16$. This is expected since large m_l dimensions are necessary to capture larger movements. Taking 16x16 blocks to perform block matching on the candidate peaks is already proven in the RS study to be a suitable value when we are dealing with HD sequences. The robust number of candidate peaks n_p is determined to be $n_p = 13$. Largely overlapping blocks are favored with a block step size tending to $s_b = 32$. Sub-pixel accuracy is

Figure 3: Illustration of multi-grid (left) and multi-scale (right) motion estimation approach. In both cases, ME is performed on higher scales at the top of the pyramid first and motion vectors are propagated down the pyramid to the lower scales by means of hierarchical candidates.

	Block width full res.	Block height full res.	mean PSNR	mean SI
62 Robust RS MEs	[8,32], [32,128]	[4,16], [16,64]	28.46	2.18
Proposed RS ME	16, 64	8, 32	28.91	2.46

Table 1: Block settings and performance in PSNR and SI of 62 robust RS MEs and the proposed RS ME. Block width and block height indicate the equivalent block sizes for the full resolution where the selected settings for the fine and course scales can be a range $[..]$ of values.

Figure 4: PSNR-Consistency trade-off plot of 6660 RS MEs.

Figure 5: Contour line of RS (red line) and PPC (blue line) MEs; several optimal MEs are highlighted.

m_l	$\{32, 64, 128\}$
m_s	$\{1, 2, 4, 8, 16\}$
n_p	$\{1, 2, 3, \dots, 20\}$
s_b	$\{16, \dots, m_l\}$
a_s	$\{0, 1\}$

Table 2: Parameter settings for PPC MEs

Figure 6: Optimal RS MEs within a limited distance from the contour line; 1745 highlighted MEs (blue) in the attractive segment.

Figure 7: PSNR-Consistency trade-off plot of 1800 PPC MEs.

Figure 8: Optimal PPC MEs within a limited distance from the contour line; 54 highlighted MEs (blue) in the attractive segment.

Figure 9: Contour line of PPC MEs derived from pixel-based SI metric with highlighted optimal MEs.

Figure 10: PSNR distances to the contour line for the best performing RS MEs within the attractive segment in PSNR and SI per sequence. The proposed ME is highlighted with a thicker blue line.

present in 61% of the 54 MEs. Compromises between computational complexity and performance are apparent when investigating motion estimation by means of PPC.

4 Discussion and Results

The proposed methodology can be applied to find robust ME settings within one ME type. When the location of the determined ME in the PSNR-Consistency trade-off graph is found, it can be compared with the location of another robust ME of a different ME type in terms of PSNR and SI performance. When the two MEs are situated sufficiently far apart (in order to avoid the effect of metric inaccuracies) a conclusive comparison is provided by the proposed PSNR-Consistency trade-off graph as is the case for e.g. the RS and the PPC MEs in Fig. 5.

We claim that the proposed methodology can find a robust ME even with suboptimal performance measures. The SI metric is evidently suboptimal in the sense that the SI output is dependent on the selected block dimension, thus the 8x8 block-based SI measure is not comparable with the

Figure 11: SI distances to the contour line for the best performing RS MEs within the attractive segment in PSNR and SI per sequence. The proposed ME is highlighted with a thicker blue line.

Figure 12: PSNR distances to the contour line for the best performing PPC MEs within the attractive segment in PSNR and SI per sequence. The proposed ME is highlighted with a thicker blue line.

Figure 13: SI distances to the contour line for the best performing PPC MEs within the attractive segment in PSNR and SI per sequence. The proposed ME is highlighted with a thicker blue line.

1x1 pixel-based SI measure. For the case of PPC, we have added a pixel-based SI evaluation (see Fig. 9), where a motion vector is assigned to each pixel in the image instead of validating the SI performance on only 8x8-pixel motion vector blocks. The range of well-performing MEs was chosen such that approximately the same number of MEs ended up in the attractive segment. When comparing the 8x8 block based SI with the pixel-based SI, we found that different MEs are returned in the attractive segment. 22% of the MEs in the attractive segment of the RS approach are not present in the attractive segment of the PPC method. Nevertheless, the histogram analysis reveals that the same ME is proposed in the case of the pixel-based SI. When observing the location of the proposed ME in Fig. 9 and Fig. 5, it is apparent that the proposed ME is located halfway between Opt. ME 3 and Opt. ME 4 in Fig. 9, whereas in Fig. 5, the same ME is located closer to Opt. ME 3, which underlines the incongruent output of the two SI metrics.

The performance of the proposed ME is analyzed and compared to other MEs within the attractive segment to determine the robustness of the chosen parameter settings. A robust ME is expected not to perform badly on any of the test sequences. Therefore, the PSNR and SI distance to the contour line are displayed in Fig. 10, Fig. 11, Fig. 12, and Fig. 13, where the proposed ME is plotted against the best MEs (in either PSNR or SI) for each sequence. For the case of the RS ME, only one ME (highlighted in black in Fig. 10 and Fig. 11) reveals on average smaller distances in both PSNR and SI. However, its SI distance from the contour line for the ‘WalkingMan’ test sequence (sequence 7 in Fig. 11) shown in the middle snapshot of the first row in Fig. 1 is clearly larger than the SI distance of the proposed ME. Hence, the proposed ME is not surpassed in robustness by any of the best MEs per test sequence. For PPC, there are three MEs with a smaller average distance in both PSNR and SI (highlighted in green, blue and black). The highest distance peaks of the three MEs are in the neighborhood of the highest distance peak of the proposed ME, indicating that the best MEs are suffering from similar performance disturbances. Future work should evaluate how the proposed ME performs compared to other metrics and other test sequences in order to come to a concluding validation of the proposed methodology.

To further confirm that the proposed methodology returns well-performing MEs which can compete with other techniques, a benchmark is provided in Table 3. An overview is given of the SI and M2SE-PSNR values of the investigated recursive-search MEs as well as the benchmark results from several methods described in literature implemented by us. These include full-search (FS) and reduced-search pattern based methods, i.e. three-step-search (TSS) [16], one-at-a-time-search (OTS) [17], diamond search (DS) [18] and hexagon-based-search (HEXBS) [19], as well as algorithms based

	Block width full res.	Block height full res.	mean PSNR	mean SI
Proposed RS ME	16, 64	8, 32	28.91	2.46
3DRS [3]	8	8	28.60	2.80
HRNM [14]	8	8	29.32	1.48
FS [15]	16	16	25.78	15.00
3SS [16]	16	16	22.80	3.90
OTS [17]	16	16	23.79	6.64
DS [18]	16	16	23.95	7.24
HEXBS [19]	16	16	23.90	7.28
MVFAST [20]	16	16	28.15	4.43
TCSBP [1]	16	16	28.31	4.07
EPZS [21]	16	16	28.69	3.77
MRST [22]	16,16,16,16	16,16,16,16	28.60	5.16
MPMVP [4]	32,16, 8, 4	32,16, 8, 4	28.13	3.80

Table 3: Performance comparison of proposed RS MEs and various techniques documented in literature. Block width and block height indicate the equivalent block sizes for the full resolution fine and course scales.

on spatio-temporal predictors, i.e. the predictive (zonal) search methods MVFAST [20] and EPZS [21], the RS methods 3DRS [3], HRNM [14] and TCSBP [1], and combined hierarchical-predictive methods, i.e. the MRST-method proposed in [22] and MPMVP from [4]. Note that the M2SE-PSNR metric favors ‘true’ motion, i.e. MEs with a better vector field consistency can outperform a full-search method. Furthermore, all methods from literature were adapted and tested with smaller block dimensions (e.g. 8x8), however, no improvement in PSNR and SI was observed.

The benchmark shows that the proposed hierarchical RS ME is outperformed solely by the sophisticated HRNM ME which employs 3-picture estimates. We conclude from these results that the proposed methodology does yield superior performing MEs among the thousands of ME parameter combinations.

5 Conclusion

In order to automatically identify parameter settings of robust MEs for retiming video sequences, a methodology is presented that can successfully deal with slightly suboptimal performance measures. A three-step approach is suggested where, firstly, the variety of conditions under which the MEs should perform well is defined and appropriate test data is selected. Secondly, the PSNR-Consistency trade-off curve marks the contour line which illustrates the achieved compromise between prediction accuracy and consistent motion vector fields. Thirdly, an attractive segment is identified which is limited to all the MEs within a defined distance from an attractive section of the contour line. Histogram analysis investigates the distribution of MEs within the attractive segment to identify the parameter settings of the MEs which are least sensitive to varying settings and thus most robust.

The developed methodology is discussed in two case studies where a design-space exploration is carried out for two different types of MEs. From 6600 hierarchical RS MEs and 1800 PPC MEs, the parameter settings of robust ME types are identified. A modification of the SI metric has returned the same parameter settings although the MEs in the attractive segment have differed by 22%, suggesting that the methodology can deal with different suboptimal performance measures. Regarding the robustness of the proposed MEs, the PSNR and SI distances from the contour lines for each sequence have been analyzed, from which it became evident that the proposed RS ME returns small average distances and is not less robust than the ME with the smallest average distances. The PPC MEs with smallest PSNR and/or SI distance per sequence suffer from similar performance disturbances as the proposed PPC ME. A conclusive validation of the methodology is needed in future work.

The methodology returns overall well-performing MEs as is shown in the benchmarking with various other techniques. A superior performance to multiple existing techniques with regard to combined PSNR/SI performance is observed.

Acknowledgements

The authors would like to thank Nico Cordes for his contribution to this work.

References

- [1] J. Wang *et al.*, “Temporal compensated motion estimation with simple block-based prediction,” *IEEE Transactions on Broadcasting*, vol. 49, pp. 241–248, Sep. 2003.
- [2] G. de Haan *et al.*, “True-motion estimation with 3-D recursive search block matching,” *IEEE Trans. Circuits*,

- Syst. Video Techn.*, pp. 368–379, Oct. 1993.
- [3] G. de Haan and P. Biezen, “Sub-pixel motion estimation with 3-d recursive search block-matching,” *Signal Processing*, vol. 6, pp. 229–239, June 1994.
 - [4] S.-C. Tai *et al.*, “A multi-pass true motion estimation scheme with motion vector propagation for frame rate up-conversion applications,” *Display Technology, Journal of*, vol. 4, pp. 188–197, June 2008.
 - [5] C. N. Cordes and G. de Haan, “Key requirements for high quality picture-rate conversion,” *SID Digest of Technical Papers*, vol. 15, pp. 850–853, June 2009.
 - [6] E. B. Bellers, “Motion compensated frame rate conversion for motion blur reduction,” *SID Digest of Technical Papers*, vol. 38, pp. 1454–1457, May 2007.
 - [7] G. A. Thomas, “Television motion measurement for datv and other applications,” Tech. Rep. PH-283, BBC Research Department, 1987.
 - [8] Snell’s Alchemist Ph.C-HD motion-compensated standards converter, see <http://www.snellgroup.com/news-and-events/press-releases/625/snell-announces-1080p-support-for-alchemist-ph.c-hd>.
 - [9] G. de Haan and E. Bellers, “De-interlacing of video data,” *IEEE Transactions on Consumer Electronics*, vol. 43, pp. 819–825, Aug. 1997.
 - [10] B. Horn and B. Schunck, “Determining optical flow,” *Artificial Intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981.
 - [11] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
 - [12] F. Dufaux and F. Moscheni, “Motion estimation techniques for digital TV: A review and a new contribution,” *Proc. IEEE*, pp. 858–876, June 1995.
 - [13] A. Heinrich *et al.*, “Optimization of hierarchical 3DRS motion estimators for picture rate conversion,” *IEEE Journal of Selected Topics in Signal Processing*, Apr. 2011. To be published.
 - [14] E. Bellers *et al.*, “Solving occlusion in frame-rate up-conversion,” in *Digest of the ICCE*, pp. 1–2, Jan. 2007.
 - [15] J. R. Jain and A. K. Jain, “Displacement measurement and its application in interframe image coding,” *IEEE Trans. Commun.*, pp. 1799–1808, Dec. 1981.
 - [16] T. Koga *et al.*, “Motion-compensated interframe coding for video conferencing,” in *Proc. Nat. Telecom. Conf.*, pp. G 5.3.1–G 5.3.5, Nov./Dec. 1981.
 - [17] R. Srinivasan and K. Rao, “Predictive coding based on efficient motion estimation,” *Communications, IEEE Transactions on*, vol. 33, pp. 888 – 896, Aug. 1985.
 - [18] S. Zhu and K. Ma, “A new diamond search algorithm for fast block-matching motion estimation,” *IEEE Transactions on Image Processing*, vol. 9, pp. 287–290, Feb. 2000.
 - [19] C. Zhu *et al.*, “Hexagon-based search pattern for fast block motion estimation,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, pp. 349–355, May 2002.
 - [20] P. Hosur and K. Ma, “Motion vector field adaptive fast motion estimation,” in *Second International Conference on Information, Communications and Signal Processing (ICICS)*, Dec. 1999.
 - [21] A. Tourapis, “Enhanced predictive zonal search for single and multiple frame motion estimation,” *Proceedings of Visual Communications and Image Processing*, pp. 1069–79, Jan. 2002.
 - [22] J. Chalidabhongse and C. Kuo, “Fast motion vector estimation using multiresolution-spatio-temporal correlations,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, pp. 477–488, June 1997.