Erschienen in: 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX). - Piscataway, NJ : IEEE, 2020. - ISBN 978-1-72815-965-2
https://dx.doi.org/10.1109/QoMEX48832.2020.9123096

Visual Quality Assessment for Interpolated Slow-motion Videos based on a Novel Database

Hui Men[1], Vlad Hosu[1], Hanhe Lin[1], Andrés Bruhn[2], Dietmar Saupe[1]

[1]Department of Computer and Information Science, University of Konstanz, Germany
[2]Institute for Visualization and Interactive Systems, University of Stuttgart, Germany
Email: {hui.3.men, vlad.hosu, hanhe.lin, dietmar.saupe}@uni-konstanz.de, bruhn@vis.uni-stuttgart.de

***Abstract*—Professional video editing tools can generate slow-motion video by interpolating frames from video recorded at a standard frame rate. Thereby the perceptual quality of such interpolated slow-motion videos strongly depends on the underlying interpolation techniques. We built a novel benchmark database that is specifically tailored for interpolated slow-motion videos (KoSMo-1k). It consists of 1,350 interpolated video sequences, from 30 different content sources, along with their subjective quality ratings from up to ten subjective comparisons per video pair. Moreover, we evaluated the performance of twelve existing full-reference (FR) image/video quality assessment (I/VQA) methods on the benchmark. In this way, we are able to show that specifically tailored quality assessment methods for interpolated slow-motion videos are needed, since the evaluated methods – despite their good performance on real-time video databases – do not give satisfying results when it comes to frame interpolation.**

***Index Terms*—visual quality assessment, slow motion, optical flow, frame interpolation**

I. Introduction

Slow-motion videos have become popular in recent years. However, not all cameras support the required frame rates at high resolutions. Video editing software, including professional ones such as "Adobe Premiere Pro CC"™, provide methods to generate slow-motion videos by synthesizing frames, starting from standard frame-rates. Thereby, the quality of the generated videos depends on the applied interpolation techniques which typically fill-in image content along the path of motion. The required motion field in the form of the so-called optical flow can be derived in several ways. Widely used approaches include block matching methods [1], frequency-based techniques [2], variational methods [3], and convolutional neural networks [4].

Computation of optical flow is a research topic on its own, and there are several benchmark datasets to evaluate and rank competing algorithms. Only one of these benchmarks also allows comparing the quality of interpolated frames: the Middlebury benchmark [5]. In this benchmark, the performance of motion estimation is evaluated by angular and endpoint errors between the estimated flow and its ground-truth. Besides, it also provides a simple objective evaluation of the corresponding motion-compensated interpolation results, given by the root mean squared error (RMSE) and the gradient normalized RMSE between the interpolated frame and the ground-truth one. However, since the Middelebury benchmark primarily aims at evaluating optical flow methods, it falls short regarding two aspects when it comes to frame interpolation. First, it only offers a small number of image triplets for interpolation, each consisting of a frame pair and the in-between ground-truth. Second, it uses objective RMSE metrics for evaluation that are known to be perceptually inaccurate [6]. For frame interpolation, this was confirmed by subjective studies on the Middlebury interpolation benchmark in [7] and [8]. Thus, simple objective measures are insufficient for evaluating interpolated frames, and this problem may be even more prevalent for interpolated video sequences. Due to the temporal variations contained in a video, inspecting a video rather than observing its constituent frames can result in different perceptual quality scores. Hence, we propose a benchmark specifically for interpolated slow-motion videos along with corresponding subjective quality scores. We also consider suitable objective evaluation metrics for comparison.

We provide 30 videos at 120 frames per second (FPS) using a high-speed camera. The whole set of videos is diverse in both content and motion types. For each original (pristine) video, we generate several slow-motion versions by interpolating subsampled frames. For this purpose, we use the same interpolation technique as adopted in the Middlebury benchmark [9] using ten optical flow methods. Playing the interpolated videos at 30 FPS makes them four times slower than their original speed. In total, our database contains 1,350 slow-motion video versions generated from the 30 source videos.

Lab studies for subjective quality assessment are well established and considered as a reliable methodology. However, the number of videos that can be assessed in the lab is limited due to the required time and cost. As an alternative, crowdsourcing studies are less expensive, and sufficiently reliable if the results are properly post-processed by removing outliers [10]. Therefore, we collect subjective scores for the slow-motion videos by crowdsourcing. Instead of using an absolute category ratings (ACR) scale, which is adopted by most of the video quality assessment (VQA) databases, we perform paired comparisons (PC) since it is a highly discriminating evaluation procedure. Moreover, instead of naively comparing

TABLE I
VQA DATABASES

Database	Year	# SRC*	# DST†	FPS	Method
EPFL-PoliMI [15]	2009	12	144	25/30	ACR
LIVE [16]	2010	10	150	25/50	ACR
IVP [17]	2011	10	128	25	ACR
CSIQ [18]	2014	12	216	24-60	ACR
CVD2014 [19]	2014	-	234	10-31	ACR
MCL-V [20]	2015	12	96	24-30	PC
NFLX [21]	2016	9	70	24-30	ACR
KoNViD-1k [13]	2017	-	1,200	30/60	ACR
FlickrVid-150k [14]	2019	-	153,841	24-120	ACR
KoSMo-1k [Ours]	2020	30	1,350	120*	PC

* source video
† distorted video
* 120 FPS in real-time. and 30 FPS for playback.

the full set of video pairs, we use a hybrid active sampling procedure [11], which further improves the efficiency of our quality assessment.

For a slow-motion video benchmark, obtaining subjective ratings for each submitted video is not practical; thus, an objective quality assessment method is needed. Since the ground-truth videos are available, we consider four full-reference (FR) VQA and eight FR image quality assessment (IQA) methods.

Our contributions can be summarized as follows:

1) We create a VQA database, KoSMo-1k, consisting of 1,350 slow-motion videos generated by ten optical flow methods.
2) We provide subjective ratings for the slow-motion videos obtained via paired comparison with active sampling. In total, 18,626 subjective ratings were collected.
3) We evaluate the performance of twelve I/VQA methods on the generated slow-motion videos, accordingly.
4) The dataset with ratings is provided in [12].

II. Related Work

Several VQA databases are available, see Table I. The sizes of these databases range from 70 to 153,841 videos. Most of these databases contain videos degraded by artificially generating distortions, in particular, compression artifacts or transmission distortions. Videos in KoNViD-1k [13] and FlickrVid-150k [14] were collected and sampled from authentic videos of different qualities. All of these databases contain only real-time videos. None of them provides slow-motion videos or distorted videos generated by frame interpolation. Regarding subjective quality scores, except for MCL-V, which adopted PC to derive subjective scores, all other databases list mean opinion scores from ACR.

Regarding the evaluation of the interpolation quality, both the Middlebury benchmark and the other datasets adopt standard metrics (e.g., MSE, PSNR, and SSIM [6]) to measure the differences between the interpolated image and the ground-truth in-between one. However, these metrics have been judged to be insufficient for evaluating interpolated frames by subjective studies on the Middlebury benchmark [7] [8].

Hybrid-MST [11] is a hybrid active sampling method aiming at aggregating scale values from a sparse PC test. After the initial round of a PC test which is randomly sampled, it actively selects the pairs for the next round based on the expected information gain (EIG) from the previous round. The pairs for the next round are indicated by the edges of the minimum spanning tree, the nodes of which are given by the videos and the edges are weighted by the inverses of the corresponding EIGs. This sampling procedure iterates until the cost reaches the test budget.

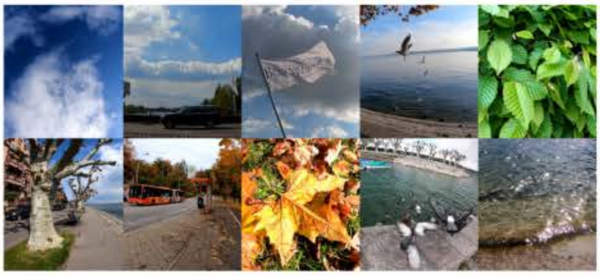


Fig. 1. Example frames from the videos in the database, sorted according to ascending bitrates. From the upper left to the lower right: 482 kbps, 598 kbps, 790 kbps, 2.14 Mbps, 2.74 Mbps, 4.01 Mbps, 4.12 Mbps, 4.74 Mbps, 5.19 Mbps and 6.50 Mbps.

III. Slow Motion Source Videos

A. General Information

Our 30 source videos were captured using a GoPro HERO7 camera at 120 FPS in MPEG-4 format (encoded with the H.265 codec). The bitrates of these source videos vary from 30 Mbps to 60 Mbps. In order to allow for subjective comparison of two slow-motion videos side-by-side, we manually scaled the videos from HD resolution of 1920×1440 to roughly half the size and cropped them to 480×540 pixels, according to content, see Fig. 3. Furthermore, since the interpolated videos are played four times slower than their original speed, i.e., at 30 FPS, we cut the videos into 2-second segments such that the slow-motion videos are 8 seconds long. The recommended video duration for subjective studies is 8–10 seconds [22]. The processed videos were stored in MPEG-4 format, encoded using the H.264 codec.

B. Video Diversity

The source videos in our database are diverse in respect of content and motion types.

1) Content Diversity: The source videos include standard scenes such as those depicting traffic, birds, landscapes, but also scenes designed to be more challenging for optical flow methods. It may be more difficult to compute flow fields for waves, clouds, sand, sparkling water (see Fig. 1). The dynamics of such scenes go beyond the usual assumption of "objects against a background".

2) Motion Diversity: The source videos are diverse in motion types as well. As shown in Table II, we classified the motion types into two main classes. One is *object motion*, meaning that the camera is relatively fixed, while the object is moving. In this class, the videos can be further grouped into three sub-classes: (i) *normal speed* (objects in the scene are moving at normal speed), (ii) *fast speed* (objects in the

TABLE II
MOTION TYPES OF SOURCE VIDEOS IN THE DATABASE

Motion Types of Source Videos		# Videos
Object Motion	Normal Speed	6
	Fast Speed	5
	Special	5
Camera Motion	Zooming	3
	Panning	3
	Tilting	2
	Dolly	4
	Trucking	3

TABLE III
SLOW-MOTION VIDEOS WITH INTERPOLATED VERSIONS

Up To*	Interp.7	Interp.15	Interp.31	Interp.63
# Source Videos	7	9	6	8
# Slow-motion videos per source	30	40	50	60
# Slow-motion videos in total	210	360	300	480

* Maximum number of frames interpolated for a source video. For example, Interp.31 indicates that from each source video slow-motion versions with 1, 3, 7, 15, and 31 interpolated frames between corresponding reference frames are included.

scene are moving fast, namely with large displacements, such as flying seagulls and fast-moving vehicles), and (iii) *special content* (the scene contains special contents, e.g., clouds, sparkling water). The other main class is *camera motion*, which can also be further subdivided according to the motion types [23]: *zooming*, *panning*, *tilting*, *dolly* and *trucking*.

IV. SLOW-MOTION VIDEO GENERATION

Before designing the interpolation strategy for generating slow-motion videos, we checked the quality differences between I-frames (which are stored completely) and other frames, i.e., non-I-frames (which are predicted from I-frames) by visually inspecting all of the frames as well as using the state-of-the-art no-reference VQA method CORNIA [24] to predict their quality scores. Both of these ways confirmed that there is no significant quality difference between I-frames and non-I-frames, thus the reference frames for the interpolation can be freely chosen.

We produced slow-motion videos from the source videos by applying ten optical flow methods with the parameters recommended by the corresponding implementations, each followed by frame interpolation from the corresponding optical flow field, using the interpolation method that had also been applied in the Middlebury benchmark, with code from [9]. See Table V for the percentile rankings of the corresponding performances in the Middlebury benchmark. We interpolated 1, 3, 7, 15, 31, and 63 frames between every other, 4th, 8th, 16th, 32nd, and 64th frame (see Fig. 2), resulting in 6 versions of slow-motion videos, denoted as *Interp-1*, *Interp-3*, *Interp-7*, *Interp-15*, *Interp-31*, and *Interp-63*). However, for some of the source videos, interpolating too many frames resulted in slow-motion videos that are severely degraded. Therefore, after visual inspection, we discarded the ones that are obviously unacceptable for viewing. This way we obtained 1,350 interpolated slow-motion videos for our database, see Table III. We treated the slow-motion videos generated from the same source video as one set. Overall there are 30 sets, each with a number of slow-motion videos according to the maximum number of frames to be interpolated.

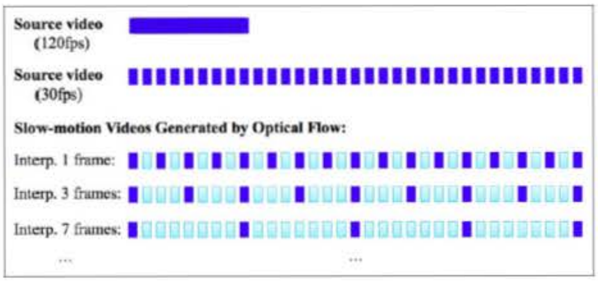


Fig. 2. Interpolation Strategy.

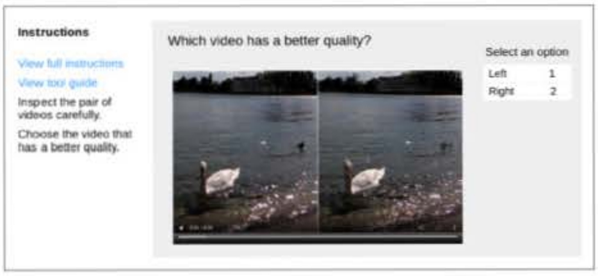
TABLE IV

	Round 1	Round 2-7	Round 8
# Pairs	3,717	4,392	1,036
# Judgements per pair	2	2	4
# Collected Judgements	16,218		4,144
# Reliable Judgements	15,052		3,574
# Reliable Judgements in total	18,626		

V. SUBJECTIVE STUDY OF SLOW-MOTION VIDEOS

A. Study Design

In order to scale the videos in each of the 30 sets according to visual quality, we collected paired comparisons (Fig. 3) using the Amazon Mechanical Turk (AMT) [25] platform. We applied the active sampling strategy (ASPC) to each of the sets in eight rounds, to avoid having to compare each video in a set to all the others. In the first round of ASPC, we randomly sampled pairs by choosing the edges of a random sparse graph with nodes corresponding to videos and a vertex degree of 6. Thus, each video is randomly compared (twice) to 6 other videos of the same set. For all sets together, this resulted in 3,717 pairs with 7,434 forced choices, see Table IV. In rounds 2 to 7, we applied the minimal spanning tree strategy of ASPC, again collecting two votes per pair. Then, we filtered outliers and re-collected the ratings for the removed pairs in the $8th$ round (4 votes for each pair). For all 8 rounds of ASPC, 8,109 pairs of videos were compared in total. After removing the outliers also for this last round, 18,626 reliable subjective ratings remained.


Instructions

Which video has a better quality?

Inspect the pair of videos carefully.

Choose the video that has a better quality.

Select an option

Left	1
Right	2

Fig. 3. Interface of crowdsourcing experiment. By clicking the play button, a pair of videos will be played simultaneously. Turkers (i.e., crowd workers working via AMT) were asked to identify and select the video with better quality for each video pair (forced binary choice). They can playback the pair of videos several times as they want.

TABLE V
PERCENTILE RANKINGS OF OPTICAL FLOW METHODS

Optical Flow Method (*Abbrev.*)	Middlebury [5]	StudyMB 2.0 [8]	KoSMo-1k
Classic+NL (*ClassicNL*) [26]	45%	21%	1
OAR-Flow (*OAR*) [27]	–	–	2
Black & Anandan (*BA*) [28] [29]	48%	26%	3
BeyondPixles (*Beyond*) [30]	–	–	4
Dual-TVL1 (*DualTVL1*) [31]	–	–	5
FFV1MT [32]	6%	24%	6
LKpyramid (*LK*) [33]	2%	6%	7
2D-CLG (*CLG*) [34] [35]	38%	42%	8
Brox et al. (*Brox*) [36] [37]	75%	88%	9
Horn & Schunck (*HS*) [38]	9%	13%	10

"–" denotes the method is not existing in the benchmark.

B. Quality Assurance

Quality control consisted of a session of thorough instructions at the beginning of the crowdsourcing tasks and later on of a step that filters outlier votes followed by their replacement by a renewed collection of PCs. During the outlier removal, individual votes were removed, based on the disagreement with the currently reconstructed score values of the presented stimuli. A vote of a crowd worker was regarded as an outlier if the worker assigned a lower score to the video with better quality and the scores of the two stimuli in the paired comparison differed by at least 0.2, which is approximatly one third of the range of the quality scale.

C. Result

Based on Thurstone's Case V model [39] with maximum likelihood estimation (code provided by [11]), absolute quality scale values for each slow-motion video were reconstructed using the judgments collected. In our study, there is no cross-content comparison, so the reconstructions in the 30 sets are independent of each other. To align the scores in all sets together, we introduced two virtual anchors per set. One stands for a slow-motion video of the worst quality among all, and the other is like the ground-truth, i.e., its quality is the best overall. After the reconstruction of the scores for each of the augmented sets, we linearly re-scaled the scale values to the interval $[0, 1]$, so that the scale values of the two anchors became 0 and 1.[1]

From these scale values we then computed an average quality for each optical flow *method* interpolating a certain *number* of frames (denoted as *method-number*[2]). Thereby the average was taken over all such interpolated sequences in all 30 sets. Fig. 4 shows these qualities depending on the number of interpolated frames for the ten considered optical flow methods. The best three performances are *DualTVL1-1*, *ClassicNL-1*, and *ClassicNL-3*; see Interp-1 and Interp-3. Fig. 5 shows the qualities depending on motion types. It can be seen that some methods performed especially well for camera motion (e.g., *OAR* and *CLG*), while some failed for most of the motion types (e.g., *HS* and *LK*).

Moreover, Table V lists the ***overall performance***, yielded by taking the average over 30 videos of the considered optical methods for KoSMo-1k. Additionally, interpolation results for the Middlebury benchmark and the *StudyMB 2.0* are listed. While the Middlebury results rely on the root-mean-square error, the *StudyMB 2.0* uses the perceptual quality of interpolated single frames. One can observe notable differences in the rankings when comparing the three cases. For instance, *Brox* ranked best for both Middlebury and the *StudyMB 2.0*, and only ninth for KoSMo-1k. In contrast, *ClassicNL* performed best on average in KoSMo-1k, but gave rather inferior results for Middlebury and the *StudyMB 2.0*.

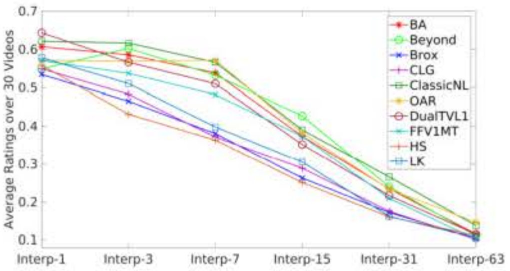


Fig. 4. Average scores over 30 videos.

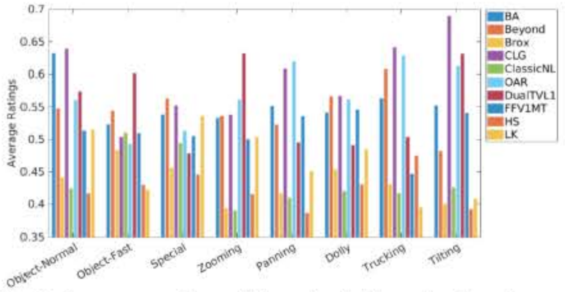


Fig. 5. Scores averaged over all interpolated videos of each motion type.

D. Discussion

As shown in Fig. 4, for two optical flow methods, the performance of interpolating three frames is better than interpolating a single frame (*Beyond* and *OAR*). While this appears to be counter-intuitive at first glance, it may be explained by the generally smoother interpolation results of those methods – which can be attributed to the recommended parameter settings in the corresponding implementations. Smoother interpolation results, in turn, may lead to slight flickering if alternated with the somewhat sharper reference frames. Not surprisingly, this effect is less pronounced, if more frames are interpolated.

In Fig. 4, there is a clear difference between two groups of methods: 1. the upper bundle, consisting of six methods that are generally better performing *BA*, *Beyond*, *ClassicNL*, *OAR*, *DualTVL1*, and *FFV1MT*; 2. the lower bundle, consisting of the other four methods that perform worse on average especially in the mid-range interpolation scenarios (Interp-3 to Interp-31), *Brox*, *CLG*, *HS*, and *LK*. The reason for the poor performance of *HS* and *LK* is that those classical methods, which have been proposed almost 40 years ago, are based on simpler assumptions that have an impact on both the accuracy and the robustness of the estimation. Regarding

[1] All reconstructed quality values, accompanied by their corresponding rankings will be shown in tables on our website.

[2] E.g., *CLG-3* denotes "optical flow method *CLG* interpolates *3* frames".

TABLE VI
SROCC OF FR-I/VQA METHODS ON DIFFERENT DATASETS

FR-VQA*	VQEG [40]	NFLX [21]	LIVE (VQA) [16]	KoSMo-1k
MOVIE [41]	0.858	–	–	0.427
ST-MAD [42]	–	–	0.824	0.443
ViS3 [43]	–	–	0.816	0.450
VMAF [44]	–	0.940	–	**0.515**

FR-IQA†	LIVE (IQA) [45]	StudyMB 2.0 [8]	KoSMo-1k
MAD [46]	0.944	0.621	0.365
PSNR [47]	0.876	0.682	0.409
GMSD [47]	0.960	0.663	0.469
VSI [48]	0.952	0.658	0.472
FSIM [49]	0.963	0.660	0.476
VIF [50]	0.964	0.422	0.476
MS-SSIM [51]	0.952	0.664	0.478
SSIM [6]	0.948	0.670	0.482

* For FR-VQA: SROCC for all datasets but KoSMo-1k were taken from their references shown in the first column.

† For FR-IQA: SROCC for LIVE-IQA were taken from their references shown in the first column; results for StudyMB 2.0 were taken from [8].

the other two methods in the lower bundle, *Brox* and *CLG*, we found that for the videos with large displacements, where most of the other methods had problems, they were able to achieve visually acceptable results. However, for the videos with normal motion, they performed much worse than other methods. Also, this could be an effect of the recommended default settings of the corresponding implementations.

VI. Evaluation of FR-I/VQA

Lastly, we investigated the performance of objective FR-I/VQA methods to predict the subjective qualities of interpolated slow-motion videos in KoSMo-1k. For this purpose, we considered twelve FR-I/VQA methods, including four for VQA (MOVIE, VMAF, ST-MAD, and ViS3) and eight for IQA (PSNR, GMSD, MS-SSIM, SSIM, VIF, MAD, FSIM, and VSI). Regarding the eight IQA methods, we applied them for each frame and took the mean as the final quality score.

Table VI shows the Spearman's rank-order correlation coefficient (SROCC) between the predictions of these FR-I/VQA methods on KoSMo-1k and several other datasets. It can be seen that all of these methods performed quite poorly on KoSMo-1k, regardless of how well they performed on other datasets or the Middlebury benchmark. More specifically, *ST-MAD* and *ViS3*, which are VQA methods, performed even worse than two IQA methods (i.e., *PSNR* and *MAD*). This means that some of the FR-VQA methods could not even predict the quality of interpolated slow-motion videos as well as frame-based FR-IQA methods. This clearly shows that existing quality assessment methods are not suitable for measuring the visual quality of interpolated slow-motion videos. Evidently, novel specifically tailored VQA methods are needed.

VII. Limitations

One limitation is the small frame size of the side-by-side videos, much smaller than during normal video consumption. Moreover, we captured the videos with a GoPro camera with a wide-angle lens. To reduce the wide-angle distortions we cropped the videos, however, in some of them there are still some wide-angle artifacts visible.

One open question that concerns the comparison of interpolation results in Fig. 4 and Table V is the selection of appropriate parameter settings for the optical flow methods. In particular, since there is no suitable measure to assess the quality of the interpolated videos, there is no obvious choice for a loss function that could be used to adjust those parameters. Hence, we resorted to those settings that are recommended in the respective implementations, which in most cases coincide with the optimal parameters for the Middlebury benchmark. But even in this case, the optimality of the parameters refers only to the quality of the flow and not the quality of the interpolated videos. Having a novel specifically tailored VQA method would resolve this problem. Then the parameters could be adjusted such that the interpolated videos provide the optimal visual experience.

VIII. Conclusion and Future Work

In this paper, we made three contributions for visual quality assessment of interpolated slow motion videos. First, we provided a novel bench-mark database specifically tailored for this task. Besides a large number of slow-motion videos interpolated with different optical flow methods, our database also offers a large variety of content and motion types. Secondly, based on this database, we provided and evaluated subjective ratings for the visual quality of the interpolated videos. Our study depicts that there are large differences in the perceptual quality of the interpolated videos generated by different optical flow methods. Finally, we evaluated the performances of current existing FR-I/VQA methods on such interpolated slow-motion videos. The poor correlations between their predictions and our subjective ratings reveal the weakness of FR-I/VQA methods when applied to slow-motion videos, generated from frame interpolation methods. In this context, some of the FR-VQA methods performed even worse than FR-IQA methods. This illustrates the need for developing an FR-VQA method that is specifically designed for interpolated slow-motion videos.

Hence, as future work, we suggest designing an FR-VQA model for the quality prediction of interpolated slow-motion videos. To this end, the 30 sets of slow-motion videos, along with their subjective ratings in KoSMo-1k, can be subdivided into subsets for training, validation, and testing which allows us to apply cross-validation using the leave-one-out strategy. Such a FR-VQA model would not only enable us to adjust the parameters of the optical flow methods to achieve optimal performance. It would also allow us to rank optical flow methods regarding their perceptual interpolation quality.

References

[1] T. Ha, S. Lee, and J. Kim, "Motion compensated frame interpolation by new block-based motion estimation algorithm," *IEEE Trans. on Consumer Electronics (TCE)*, vol. 50, no. 2, pp. 752–759, 2004.

[2] S. Meyer, O. Wang, H. Zimmer, M. Grosse, and A. Sorkine-Hornung, "Phase-based frame interpolation for video," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1410–1418.

[3] L. L. Rakêt, L. Roholm, A. Bruhn, and J. Weickert, "Motion compensated frame interpolation with a symmetric optical flow constraint," in *International Symposium on Visual Computing (ISVC)*, 2012, pp. 447–457.

[4] H. Jiang, D. Sun, V. Jampani, M.-H. Yang, E. Learned-Miller, and J. Kautz, "Super slomo: High quality estimation of multiple intermediate frames for video interpolation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 9000–9008.

[5] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. J. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," *International Journal of Computer Vision (IJCV)*, vol. 92, no. 1, pp. 1–31, 2011.

[6] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. on Image Processing (TIP)*, vol. 13, no. 4, pp. 600–612, 2004.

[7] H. Men, H. Lin, V. Hosu, D. Maurer, A. Bruhn, and D. Saupe, "Visual quality assessment for motion compensated frame interpolation," in *International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2019, pp. 1–6.

[8] H. Men, V. Hosu, H. Lin, A. Bruhn, and D. Saupe, "Subjective annotation for a frame interpolation benchmark using artifact amplification," *arXiv e-prints*, p. arXiv:2001.06409, Jan 2020.

[9] https://github.com/Megamusz/frame-interpolation.

[10] F. Ribeiro, D. Florencio, and V. Nascimento, "Crowdsourcing subjective image quality evaluation," in *2011 18th IEEE International Conference on Image Processing*. IEEE, 2011, pp. 3097–3100.

[11] J. Li, R. Mantiuk, J. Wang, S. Ling, and P. Le Callet, "Hybrid-MST: A hybrid active sampling strategy for pairwise preference aggregation," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2018, pp. 3475–3485.

[12] http://database.mmsp-kn.de.

[13] V. Hosu, F. Hahn, M. Jenadeleh, H. Lin, H. Men, T. Szirányi, S. Li, and D. Saupe, "The Konstanz natural video database (KoNViD-1k)," in *International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2017, pp. 1–6.

[14] F. Götz-Hahn, V. Hosu, H. Lin, and D. Saupe, "No-reference video quality assessment using multi-level spatially pooled features," *arXiv preprint arXiv:1912.07966*, 2019.

[15] F. D. Simone, M. Tagliasacchi, M. Naccari, S. Tubaro, and T. Ebrahimi, "A H.264/AVC video database for the evaluation of quality metrics," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2010, pp. 2430–2433.

[16] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Trans. on Image Processing*, vol. 19, no. 6, pp. 1427–1441, 2010.

[17] F. Zhang, S. Li, L. Ma, Y. C. Wong, and K. N. Ngan, "IVP Subjective Quality Video Database," The Chinese University of Hong Kong, http://ivp.ee.cuhk.edu.hk/research/database/subjective/, 2011.

[18] P. V. Vu and D. M. Chandler, "ViS3: An algorithm for video quality assessment via analysis of spatial and spatiotemporal slices," *Journal of Electronic Imaging (JEI)*, vol. 23, no. 1, pp. 013 016–013 016, 2014.

[19] M. Nuutinen, T. Virtanen, M. Vaahteranoksa, T. Vuori, P. Oittinen, and J. Häkkinen, "CVD2014—a database for evaluating no-reference video quality assessment algorithms," *IEEE Trans. on Image Processing*, vol. 25, no. 7, pp. 3073–3086, 2016.

[20] J. Y. Lin, R. Song, C.-H. Wu, T. Liu, H. Wang, and C.-C. J. Kuo, "MCL-V: A streaming video quality assessment database," *Journal of Visual Communication and Image Representation*, vol. 30, pp. 1–9, 2015.

[21] "NFLX Public Dataset," https://github.com/Netflix/vmaf/blob/master/resource/doc/datasets.md.

[22] P. ITU-T Recommendation, "Methods for the subjective assessment of video quality, audio quality and audiovisual quality of internet video and distribution quality television in any environment," International Telecommunication Union, Tech. Rep., 2008. [Online]. Available: https://www.itu.int/rec/T-REC-P.913-201603-I

[23] https://blog.storyblocks.com/video-tutorials/7-basic-camera-movements/.

[24] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 1098–1105.

[25] Amazon Mechanical Turk (AMTurk) , https://www.mturk.com/.

[26] D. Sun, S. Roth, and M. J. Black, "Secrets of optical flow estimation and their principles," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 2432–2439.

[27] D. Maurer, M. Stoll, and A. Bruhn, "Order-adaptive and illumination-aware variational optical flow refinement." in *British Machine Vision Conference (BMVC)*, 2017, pp. 1–13.

[28] M. J. Black and P. Anandan, "The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields," *Computer Vision and Image Understanding*, vol. 63, no. 1, pp. 75–104, 1996.

[29] "A modern matlab implementation of the Black & Anandan method by Deqing Sun," http://cs.brown.edu/ dqsun/research/software.html.

[30] C. Liu, "Beyond pixels: exploring new representations and applications for motion analysis," Ph.D. dissertation, Massachusetts Institute of Technology, 2009.

[31] C. Zach, T. Pock, and H. Bischof, "A duality based approach for realtime TV-L 1 optical flow," in *German Conference on Pattern Recognition*, 2007, pp. 214–223.

[32] F. Solari, M. Chessa, N. K. Medathati, and P. Kornprobst, "What can we expect from a V1-MT feedforward architecture for optical flow estimation?" *Signal Processing: Image Communication*, vol. 39, pp. 342–354, 2015.

[33] J.-Y. Bouguet, "Pyramidal implementation of the affine Lucas Kanade feature tracker: Description of the algorithm," *Intel Corporation*, vol. 5, no. 1-10, p. 4, 2001.

[34] A. Bruhn, J. Weickert, and C. Schnörr, "Lucas / Kanade meets Horn / Schunck: Combining local and global optic flow methods," *International Journal of Computer Vision (IJCV)*, vol. 61, no. 3, pp. 211–231, 2005.

[35] J. Jara-Wilde, M. Cerda, J. Delpiano, and S. Härtel, "An implementation of combined local-global optical flow," *Image Processing On Line*, vol. 5, pp. 139–158, 2015.

[36] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in *European Conference on Computer Vision (ECCV)*, 2004, pp. 25–36.

[37] https://de.mathworks.com/matlabcentral/fileexchange/17500-high-accuracy-optical-flow.

[38] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence (AI)*, vol. 17, no. 1-3, pp. 185–203, 1981.

[39] L. L. Thurstone, "A law of comparative judgment." *Psychological Review*, vol. 34, no. 4, p. 273, 1927.

[40] "FRTVPhase I," https://www.its.bldrdoc.gov/vqeg/projects/frtv-phase-i/frtv-phase-i.aspx.

[41] K. Seshadrinathan and A. C. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *IEEE Trans. on Image Processing (TIP)*, vol. 19, no. 2, pp. 335–350, 2010.

[42] P. Vu, C. Vu, and D. Chandler, "A spatiotemporal most-apparent-distortion model for video quality assessment," in *IEEE International Conference on Image Processing (ICIP)*, 2011, pp. 2505–2508.

[43] P. V. Vu and D. M. Chandler, "Vis3: an algorithm for video quality assessment via analysis of spatial and spatiotemporal slices," *Journal of Electronic Imaging (JEI)*, vol. 23, no. 1, p. 013016, 2014.

[44] Netflix, "Toward a practical perceptual video quality metric," 2016, https://medium.com/netflix-techblog/toward-a-practical-perceptual-video-quality-metric-653f208b9652.

[45] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. on Image Processing (TIP)*, vol. 15, no. 11, pp. 3440–3451, 2006.

[46] E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging (JEI)*, vol. 19, no. 1, p. 011006, 2010.

[47] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Trans. on Image Processing (TIP)*, vol. 23, no. 2, pp. 684–695, 2014.

[48] L. Zhang, Y. Shen, and H. Li, "VSI: A visual saliency-induced index for perceptual image quality assessment," *IEEE Trans. on Image Processing (TIP)*, vol. 23, no. 10, pp. 4270–4281, 2014.

[49] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: a feature similarity index for image quality assessment," *IEEE Trans. on Image Processing (TIP)*, vol. 20, no. 8, pp. 2378–2386, 2011.

[50] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 3. IEEE, 2004, pp. 7009–712.

[51] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *IEEE Asilomar Conference on Signals, Systems & Computers*, vol. 2, 2003, pp. 1398–1402.