No-Reference Error-Tolerability Evaluation for Videos via Edge and Extreme-Value Checking

Tong-Yu Hsieh, Shang-En Chan, Chao-Ru Chen, Pao-Chien Li and Chi-Hsuan Ho
Department of Electrical Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan

Abstract—**Approximate computing has been shown to be a promising manner to enhance circuit performance via acceptable quality degradation. This manner is applicable to many multimedia applications such as videos. In order to carry out a dynamic approximate computing scheme, on-line quality evaluation/monitoring will be needed. However, to the best of our knowledge, on-line error-tolerability evaluation of videos is seldom studied in the literature. To achieve this, a no-reference manner will be helpful, which means that no reference videos are needed for comparison with the videos under test. Implementation of on-line test procedures can thus be greatly simplified. In this paper we investigate how to achieve no-reference error-tolerability evaluation for videos where the background may be fixed or changed. We will show that by well exploiting some simple attributes, the acceptability for 81,412 erroneous videos with fixed background can be accurately determined with more than 90% accuracy. As a comparison, the related previous work can only achieve about 80% accuracy. In addition, our attribute acquirement process requires only 33% of the computation time for the previous work. As for videos with background changing, we also discuss possible solutions.**

I. INTRODUCTION

In IoT (Internet-of-Things) applications such as smart cities or smart automotive [1], video processing circuits are extensively utilized to identify objects of interests such as cars or pedestrians. For successful identification, the integrity of the video data is crucial. However, errors usually appear during capture, storage or processing of videos due to aging effects of cameras, memories or video processing circuits. These may significantly degrade the quality of the videos and even invalidate the systems.

Fortunately, video errors would not always induce drastic failures unless the structure of the object to be identified is significantly destroyed. This notion is referred to as *error-tolerance* [2], which can relax correctness requirement of video data. As a result, the lifetime of the video-based systems can be extended. This concept has great applicability to many multimedia applications as illustrated in [3]-[5]. In this work, we focus on video applications.

For video-based IoT applications, video processing is usually performed in a real-time way. This makes evaluating error-tolerability of the video related circuits such as memories and video processing circuits a great challenge. Several major issues thus arise, listed below.

1. How to carry out error-tolerability testing of videos in an ***on-line*** manner?
2. How ***accurate*** can the on-line error-tolerability testing be achieved? In other words, how much is the probability that an acceptable (unacceptable) video will be determined to be acceptable (unacceptable) after testing?
3. How much ***cost*** will be incurred to perform on-line error-tolerability testing?

In the literature there have been a number of accurate error-tolerability test methods developed [6]-[8]. However, these methods focus on the *off-line manufacturing test* scheme [9] that needs to stop the normal function of the target system and performs test procedures in the test mode. This, however, may not be allowed in real-time systems such as automotive.

Although there are typical on-line test methods such as parity check or ECC [9], *no on-line test methods so far can support error-tolerability testing* to the best of our knowledge. Also large area and latency cost may be incurred by these methods, and their test effectiveness may be limited as well. For example, at most two erroneous bits may be detected.

In this work we investigate these issues. A no-reference error-tolerability test scheme is developed. The experimental results for the 81,412 erroneous videos show that 90.44%~91.81% accuracy can thus be achieved. As a comparison, related previous work can achieve only 77.27%~83.56% accuracy. We also compare the required computation complexity to acquire the identified attributes with the related previous work. The comparison result shows that our acquirement process requires only 33% of the computation time for the previous work. Some preliminary results of this work can be found in [10]. In this paper we address much more detailed issues. In addition, we also present discussions on test solutions for videos where the background content may be changed in some time intervals.

In the rest of this paper we first review related previous work in Section II. Section III then describes how we generate erroneous videos. The error analysis and error-tolerability test scheme is presented in Section IV. Section V shows the experimental results and comparisons. Solutions to deal with videos where the background may be changed are given in Section VI. Finally, this paper is concluded in Section VII.

II. PREVIOUS NO-REFERENCE IMAGE QUALITY ASSESSMENT METHODS

A video is usually composed of a series of frames (images). In general a smooth video requires at least 15 frames outputted per second. In the literature there have been several no-reference image quality assessment methods developed. In these methods a certain particular attribute of an image (e.g., a frame in a video) is usually identified such as blurring and blocking. More importantly, such attributes can be acquired based on only the content of the target image without the reference image. One representative example of the previous no-reference methods is to evaluate the sharpness of an image [11]. Image sharpness may be degraded due to the blurring effect that usually occurs during the video acquisition, processing or compression. In [11] the authors estimate the detection probability of blurring in the image based on a probabilistic model. In particular, the sensitivity of human beings to blurring at different contrasts is considered in this model. In the experimental results, the superior performance of the method developed in [11] has also been shown by comparing with the related previous work.

For the previous no-reference methods, they mainly target a single specific quality distortion due to transmission or compression of the video data. A certain specific attribute can thus be identified to model and quantify the distortion. Erroneous images due to defective circuits (e.g., caused by aging), however, contain quite different types of distortions in addition to those considered in the previous methods. As a result, the previous methods would not be able to deal with videos with such images (frames). Later in this paper we will elaborate this issue and illustrate the inefficacy of the previous methods by using the work of [11].

Another issue for the previous no-reference methods is that *high computation cost* will be required. The main reason is that *a fine and detailed grading score* is usually provided for these methods so as to objectively determine which level of quality the target video has. Nevertheless, based on error-tolerance, *binary classification* is sufficient, i.e., we only need to know if the error is acceptable or not. In this paper we will show that this consideration will be very helpful to simplify the video quality evaluation process.

Fig. 1: Employed benchmark videos

III. Generation of Erroneous Videos

A. Employed Benchmark Videos

A total of four benchmark videos are employed in this work, which are shown in Fig. 1, each of which contains 300 frames (10 seconds) with 176*144 (=25,344) pixels per frame. These videos are commonly used in the multimedia field to benchmark video processing methods.

B. Error Injection

To facilitate generation of erroneous videos, we employ the open source NOVA H.264 decoder hardware [12]. 20,353 single stuck-at faults are injected in each component of the decoder hardware. The employed benchmark videos are then applied to each faulty decoder, and a total of 81,412 erroneous videos are thus generated.

In this work, we use the attribute of SSIM as a golden model for quantifying the acceptability of the generated videos and for comparison with the evaluation result by the developed test scheme. SSIM [13] is one of the most accurate video quality evaluation attributes developed in the literature. In the case that an erroneous video is more similar to the error-free one, the SSIM value will be higher. We find that **0.9** can be an acceptable SSIM threshold.

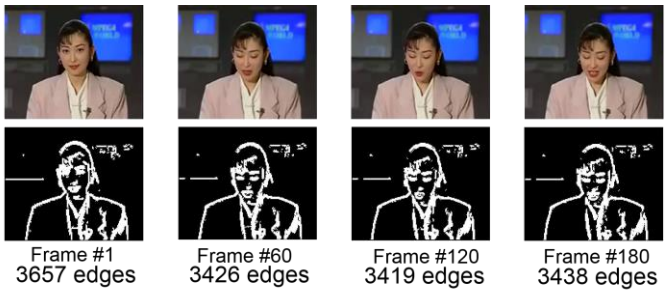

Fig. 2: Total number of edges for an error-free video

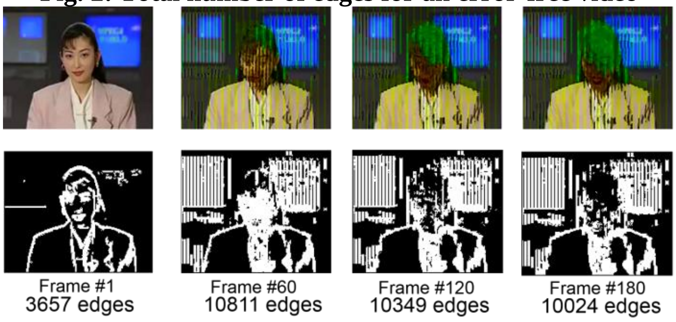

Fig. 3: Total number of edges for an unacceptable video

IV. Characteristics Analysis of Erroneous Videos and Its Application to Error-Tolerability Test

A. Variance in Number of Edge Pixels between Video Frames

By carefully analyzing the generated videos, we find that one critical attribute that can reflect the structural variance is *the difference* between the *total number of edge pixels of the first frame and that of other frames* of the video. This difference will increase more significantly for an unacceptable video than that for an acceptable video. This is mainly because that the first decoded video frame is usually employed as feedback by the faulty decoder iteratively to decode other frames, and thus the errors will be accumulated with the decoding of each frame.

Table 1: Values of $Edge_{diff}$ for error-free videos

Benchmark	$Edge_{diff}$
Akiyo	4.2%
Container	1.82%
Hall	2.6%
Silent	2.33%

In Fig. 2 and Fig. 3 we show the number of edge pixels for the error-free video and an unacceptable video (SSIM=0.4861), respectively. Four representative frames, namely the 1st, 60th,

120^{th} and 180^{th} frame, are considered for the illustration purpose.

As illustrated in Fig. 2 the total numbers of edge pixels in the frames are similar for an error-free video. Without loss of generality, in Table 1 we show the averaged difference of total number of edge pixels between the first frame and the others for all the four error-free benchmark videos. Throughout this paper we refer to this difference as $\boldsymbol{Edge_{diff}}$. As can be seen, $Edge_{diff}$ ranges from 1.82% to 4.2% in the error-free video.

For unacceptable videos, as shown in Fig. 3, *much more false edges* appear due to the errors. As a result, significant $Edge_{diff}$ appears (larger than 174%). In Fig. 4 we show the relationship between the video quality (SSIM) and $Edge_{diff}$ for the generated erroneous Akiyo videos. As can be seen, basically when the video quality is still good (SSIM $\geq$ 0.9), $Edge_{diff}$ is small. On the other hand, $Edge_{diff}$ becomes relatively large in most cases when the video quality is unacceptable. Similar results are also found for the other three benchmark videos. One issue is that unacceptable videos and acceptable ones may have similar $Edge_{diff}$ in some cases as shown in Fig. 4. As a result, it is not a trivial task to classify acceptability of erroneous videos based on their $Edge_{diff}$ values. Another issue is that different benchmark videos may have different SSIM-$Edge_{diff}$ relationship. The identified classification criterion should be generally applicable to different videos for the on-line test concern.

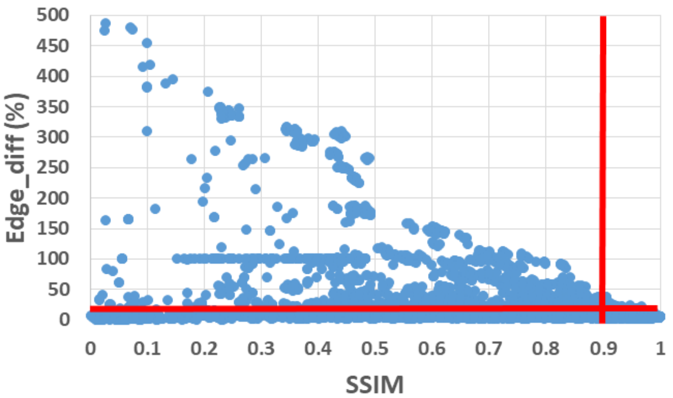

Fig. 4: Relationship between SSIM and $\boldsymbol{Edge_{diff}}$

To address the issues mentioned above, we perform a detailed **numerical analysis** on possible $Edge_{diff}$ thresholds. All the generated erroneous videos for the four benchmark videos are also considered together to perform an extensive analysis. Interestingly we find that a proper upper bound and a lower bound can be identified to accurately classify acceptability of videos for all the considered videos. As illustrated in Fig. 4, a proper upper bound is **10%**, which is indicated by the red horizontal line. We find that for the Akiyo benchmark, there is an 88.5% probability that the $Edge_{diff}$ for unacceptable (acceptable) erroneous videos is larger than (smaller than or equal to) 10%. On the other hand, we also find that when the quality of the video is quite bad (e.g., SSIM <0.3), $Edge_{diff}$ will also become quite small (e.g., <**1%**). This is because when the error is very significant, most or even all the information of the video may be lost. As a result, 1% can be a lower bound where the video whose $Edge_{diff}$ is smaller than 1% will be classified to be unacceptable. By taking this lower bound into consideration together, there is an 89.27% probability that the acceptability of an erroneous video can be correctly classified for Akiyo. Similar accuracy results are also found for the other considered benchmark videos.

B. *Number of Extrem-Value Pixels in Video Frames*

In order to further increase the probability of correct classification, we carefully analyze the misclassified videos by $Edge_{diff}$. We find that in some cases the unacceptable videos may have a similar number of edge pixels to that of acceptable ones, but a large number of pixels have *extreme values* (close to 0 or 255). By analyzing such videos, we find that many pixels of the frames in the video have the value either larger than 240 or smaller than 40. Based on these thresholds we determine if a pixel has an extreme value. Accordingly the total number of extreme-value pixels in each frame is normalized by the total number of pixels in the frame. The normalized result is then averaged, and referred to as $\boldsymbol{N_{extreme}}$.

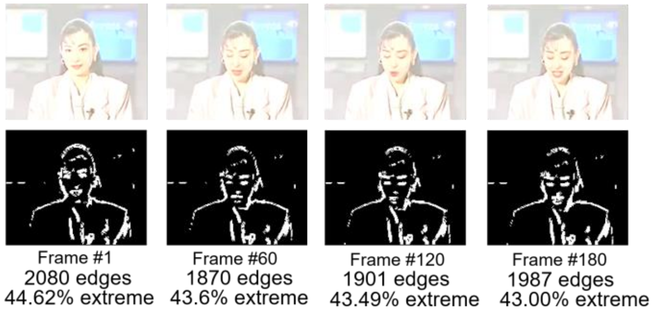

Fig. 5: Example of the extreme-value issue

One example of the extreme-value issue is illustrated in Fig. 5 where an erroneous video with the SSIM of 0.4833 is considered. As can be seen, there is only minor difference between the numbers of edge pixels in these frames, but the fraction of extreme-value pixels these frames contain is over 43%. Note that the brightening problem shown in Fig. 5 is just one example for the extreme-value issue. This issue may also result in other types of unacceptable erroneous videos. By *additionally examining the total number of pixels that contain extreme values*, we can identify unacceptable videos resulted from the extreme-value issue.

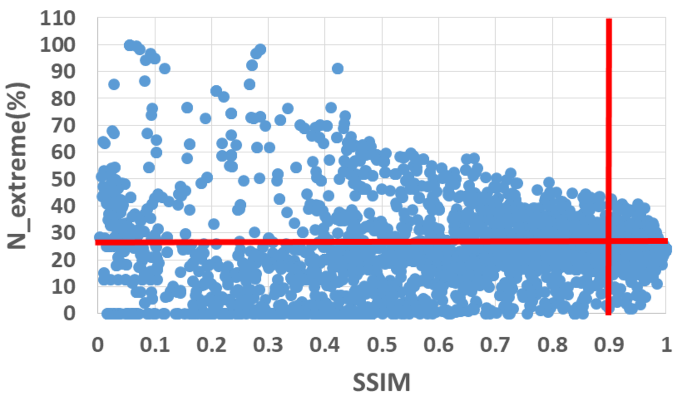

Fig. 8: Relationship between SSIM and $\boldsymbol{N_{extreme}}$

In Table 2 we show the $N_{extreme}$ values for all the considered error-free benchmark videos. As can be seen, significant variances exist for these videos, ranges from 7.44% to 23.98%, which depends on the color characteristics of these videos. This, however, complicates the determination of proper classification thresholds based on $N_{extreme}$. In Fig. 8 we show the relationship between the video quality (SSIM) and $N_{extreme}$ for the generated erroneous Akiyo videos. As can be seen, when $N_{extreme}$ is large (e.g., >50%), it is quite likely that the quality of the erroneous video is unacceptable. However, this threshold value would also induce mis-classification of many unacceptable videos to be acceptable. Based on extensive analysis, we find that a proper upper bound is **28%** that can globally achieve a good classification probability for all the considered benchmark videos. This upper bound is shown by the red horizontal line in Fig. 8. It should be noted that there also exists a lower bound on $N_{extreme}$, where the video whose $N_{extreme}$ is smaller than **1.5%** will have unacceptable quality. By considering these thresholds together, there is an 83.54% probability that the acceptability of an erroneous video can be correctly classified by $N_{extreme}$ for Akiyo. Similar accuracy results are also found for the other considered benchmark videos.

Table 2: Values of $N_{extreme}$ for error-free videos

Benchmark	$N_{extreme}$
Akiyo	23.98%
Container	7.44%
Hall	10.09%
Silent	11.34%

C. *Application to Error-Toleraiblity Test*

As shown in the previous two sub-sections, both $Edge_{diff}$ and $N_{extreme}$ can contribute high accuracy. We thus classify the acceptability of a video based on $Edge_{diff}$ and $N_{extreme}$ as follows.

By checking $Edge_{diff}$ and $N_{extreme}$ of a target video, if both the two tests are passed (i.e., $1\% \leq Edge_{diff} < 10\%$ and $1.5\% \leq N_{extreme} < 28\%$), then the video is determined to be acceptable. Otherwise, it is an unacceptable video.

V. EXPERIMENTAL RESULTS

In order to evaluate the classification accuracy of this test scheme, we apply this scheme to the generated 81,412 erroneous videos. If the acceptability classification result by this scheme matches that by SSIM, we say the test result is accurate. Accordingly we calculate the fraction of accurate test results for the generated erroneous videos as the **test accuracy** of the developed test scheme.

The test accuracy evaluation results are shown in Fig. 10 where we compare the test accuracy by (1) only $Edge_{diff}$, (2) $Edge_{diff}$ and $N_{extreme}$, and the previous no-reference video quality evaluation method [11]. As can be seen, by considering only $Edge_{diff}$, the test accuracy ranges from 88.2% to 89.81%. Considering both $Edge_{diff}$ and $N_{extreme}$ enhances the test accuracy to 90.44%~91.81%. Also only a small variance appears for these test accuracies, showing the general applicability of the developed test scheme. As for [11], only 77.27%~83.56% test accuracy is achieved.

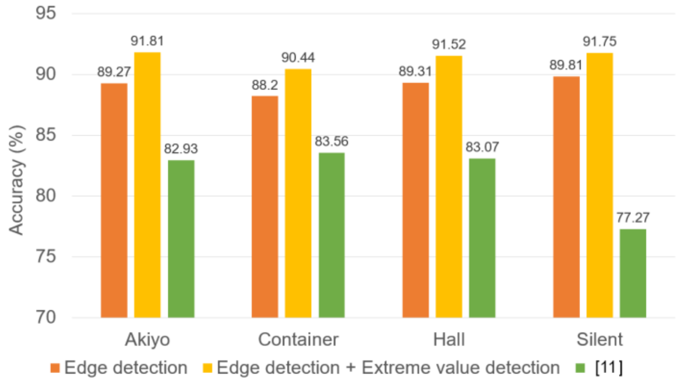

Fig. 10: Test accuracy comparison

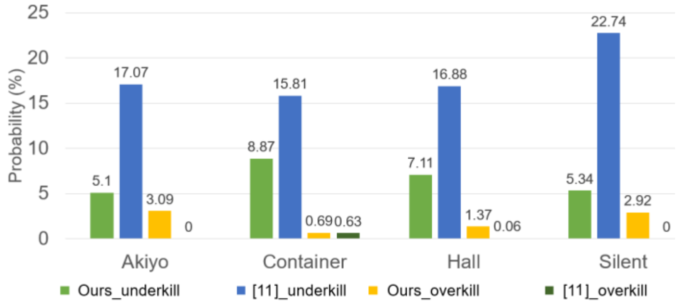

Fig. 11: Analysis of mis-classification

Furthermore, we also analyze mis-classifications by the developed test scheme and the evaluation method in [11]. The occurrence rates of *overkill* (i.e., the video is acceptable but fails our test) and *underkill* (i.e., the video is unacceptable but passes our test) are investigated, and the result is shown in Fig. 11. As indicated, the occurrence rate of under-kill for our developed test scheme ranges from 5.1% to 8.87%, which is actually much smaller than that for the method in [11]. We find that this problem is mainly caused by the fact that in some cases the first video frame has been already damaged significantly (but $N_{extreme}$ is still acceptable). In such cases, although the later frames become worse, $Edge_{diff}$ is not large/small enough to be unacceptable. As a result, the SSIM is unacceptably low, but the video would pass our test. In order to address this issue, one possible solution is to develop an accurate no-reference method to examine the quality of the first video frame. We believe this will be a quite valuable extension of this work, and is our on-going research. As for overkill, the occurrence rate for both our test scheme and [11] is small.

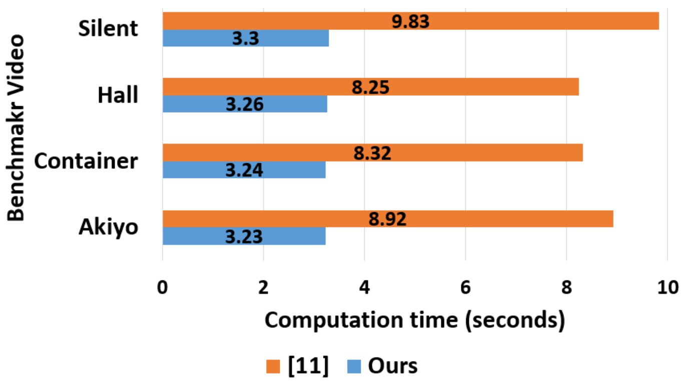

Fig. 12: Computation time comparison

In Fig. 12 we compare the required computation time for our test scheme and the method in [11] to acquire attribute values. Both the two methods are carried out using the Matlab software and run at a machine with a 2.4GHz processor and 80GB memory. As indicated, our test scheme requires only 33% of the computation time for [11]. This is mainly because that we employ much simpler attributes in our test scheme, which greatly simplifies the computation process.

VI. DISCUSSIONS

As shown in the experimental results above, the proposed test scheme works well for videos where the background content is fixed. This case usually appears for surveillance-based video applications. In this section we discuss a different case that video background may be changed in different time intervals. This case usually appears for ADAS (Advanced Driver Assistance Systems) for automotive such as pedestrian detection, lane departure, etc.

One possible solution to deal with this case is that the video can be first divided into a number of scenes. Then $Edge_{diff}$ examination can be applied to each identified scene. In the literature there also have been a number of video scene detection works that can perform this task [14]. In this way, the proposed scheme is still applicable. One major concern for this solution is that additional computation time is required for scene detection before applying the proposed test scheme. As a result, whether the whole acceptability examination process can be done in real-time still needs further investigation.

In our on-going research, we are working on developing solutions that can carefully examine the acceptability of each single video frame and accordingly determine the acceptability of the whole video. Different from the $Edge_{diff}$ based test scheme that relies on comparison with the first referenced frame, now we can monitor the acceptability of each decoded frame without any reference frames. The basic idea of this solution is explained in the following.

We find that the total number of “**false edges**” can be used as an attribute to effectively quantify significance of an erroneous image. When the error is more significant, *large deviation* would appear in the pixel values, which would lead to more *edges* shown in the image. However, some of these edges actually do not exist in the error-free images. We thus refer to such edges as *false edges*. Based on this observation, an attractive no-reference error-tolerability test methodology is developed. This methodology first identifies edge pixels in the target image, and accordingly examine total number of false-edge pixels. If this number is larger than the acceptable threshold, the target image is determined to be unacceptable. Otherwise, the number of pixels that have extreme-values will be further examined. When the target image passes both the false-edge pixel and extreme-value pixel tests, the image is determined to be acceptable. We have employed 126,894 images to generally evaluate the effectiveness of the proposed test methodology. The experimental results show that up to 93.39% test accuracy is achieved on average by the proposed methodology.

We find that for unacceptable erroneous images, the structure of the image content tends to be significantly modified. Interestingly, we also find that the *total number of false-edge pixels* (i.e., those edge pixels that do not appear in the error-free image) will increase more significantly than that of an acceptable image. This is illustrated in Fig. 13 where an acceptable image and an unacceptable image are considered, respectively. In Fig. 13 the edges in an image are emphasized by using a binarized manner. That is, if a pixel in an image belongs to an edge, its corresponding value in the binarized image equals 255 (i.e., a white pixel). Otherwise it equals 0 (i.e., a black pixel). Note that 255 (0) is the maximum (minimum) possible value of a pixel. As can be seen, the identified edges of the acceptable image is almost the same as those of the error-free image. The fraction of false-edge pixels is 2.12%. As for the unacceptable image, the more significant errors result in significant modification on the structure of the image, and thereby many false-edge pixels appear, as illustrated by the red rectangles. The fraction of false-edge pixels is increased to 17.61%. According to our preliminary study in [15], for an acceptable image, the fraction of false edges is below 5.4%. On the other hand, the fraction of false edges for most of the unacceptable images is larger than 5.4%. As a result, 5.4% can be a threshold for acceptability testing on Lena images. It should be noted that we also find that there exist different best thresholds for different images for achieving the best test accuracy. For example, for high-frequency images such as Baboon where there are more real-edge pixels, a higher threshold (e.g., 9.34%) should be employed. We thus develop a *dynamic* threshold determination method that can adaptively select a proper threshold according to the frequency of the target image. More details about implementations of this methodology can be found in [15].

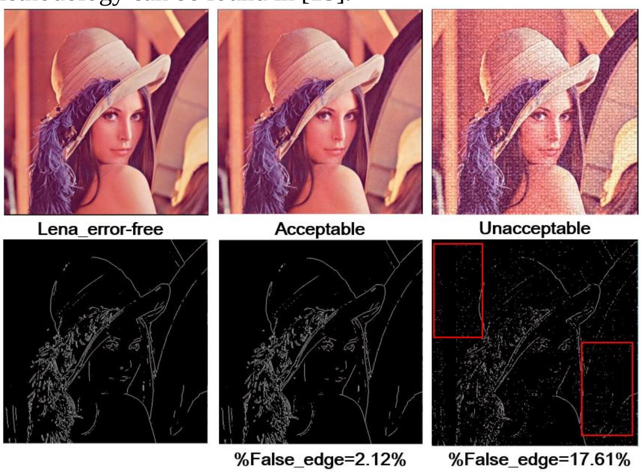

Fig. 13: Fraction of false edges for erroneous images

Fig. 14 shows the experimental results when applying both the $Edge_{diff}$ based test scheme and the false-edge based test methodology to an erroneous yet acceptable video. Here we employ the commonly used benchmark video “foreman” where there exist changes in the background content. In Fig. 15 we show some representative frames with background changing. As can be seen in Fig. 14, for the $Edge_{diff}$ based test scheme, some significant deviation on $Edge_{diff}$ may appear due to background changing as indicated by the red rectangle. Since the video frame content after background changing is quite different with the first reference frame, large $Edge_{diff}$ is resulted and thus the video is determined to be unacceptable (acceptability misclassification). On the other hand, for the false-edge based test methodology, the acceptability of each frame is evaluated independent of any reference frame. As a result, the detection result for frames is much more consistent. Therefore the acceptability of the video is accurately determined. This illustrates the great potential of the false-edge based test methodology in implementing accurate on-line error-tolerability testing for videos.

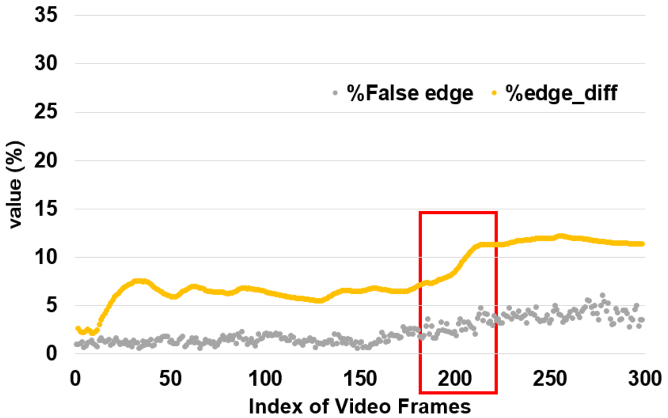

Fig. 14: Experimental results for Foreman video

Fig. 15: Background changing in Foreman video

Much more work can still be done to optimize the performance and cost of the test process discussed above. Some examples are described as follows, which are also our on-going research directions.

- Test accuracy evaluation for a larger set of videos, including those specifically for automotive applications.
- Test accuracy enhancement of the $Edge_{diff}$ based test scheme and the false-edge based test methodology.
- Integration of the $Edge_{diff}$ based test scheme and the false-edge based test methodology to carry out a more efficient test process.
- Hardware design and optimization of the test architecture.
- Consideration of transient errors.

VII. Conclusions

In this work we have investigated the development of an efficient no-reference on-line error-tolerability test scheme for videos. Our result shows that by well analyzing the target video itself, much higher (over 90%) test accuracy can be achieved when compared with the previous work, and the required computation time can be reduced significantly as well.

Acknowledgement: This work was supported in part by the Ministry of Science and Technology of Taiwan under Contract Number MOST 103-2221-E-110-077-MY3, MOST 106-2218-E-110-007, and MOST 106-2221-E-110-075.

References

[1] J. Gubbi, R. Buyya, S. Marusic and M. Palaniswami, “Internet of Things (IoT): A vision, architectural elements, and future directions,” Future Generation Computer Systems, 29(7): pp. 1645-1660, 2013.
[2] M. A. Breuer, S. K. Gupta and T. M. Mak, “Defect and error-tolerance in the presence of massive numbers of defects,” IEEE Design & Test of Computers, 21(3): pp. 216-227, 2004.
[3] H.-M. Chang, J.-L. Huang, D.-M. Kwai, K.-T. Cheng and C.-W. Wu, "Low-cost error tolerance scheme for 3-D CMOS imagers," IEEE Trans. on Very Large Scale Integration, 21(3): pp. 465-474, 2013.
[4] M. A. Breuer and, H. Zhu “An illustrated method for analysis of error tolerance,” IEEE Design & Test of Computers, 25(2): pp. 168-177, 2008.
[5] D. Nowroth, I. Polian and B. Becker, “A study of cognitive resilience in a JPEG compressor,” Proc. IEEE Int’l. Conf. on Dependable Systems and Networks, pp. 32-41, 2008.
[6] Z. Pan and M. A. Breuer, “Estimating error-rate in defective logic using signature analysis,” IEEE Trans. on Computers, 56(5): pp. 650-661, 2007.
[7] S. Shahidi and S. K. Gupta, “Estimating error rate during self-test via one’s counting,” Proc. Int’l. Test Conf., pp. 1-9, 2006.
[8] T.-Y. Hsieh, K.-J. Lee and M. A. Breuer, "An error rate based test method to support error-tolerance," IEEE Trans. on Reliability, 57(1): pp. 204-214, 2008.
[9] L.-T. Wang, C. E. Stroud, and N. A. Touba, *System on Chip Test Architectures*, Morgan Kaufmann, 2008.
[10] T.-Y. Hsieh, S.-E. Chan and C.-H. Ho, "On no-reference on-line error-tolerability testing for videos," To be presented at IEEE European Test Symp., pp. 1-2, May 2018.
[11] N. D. Narvekar and L. J. Karam, "A no-reference image blur metric based on the cumulative probability of blur detection (CPBD)," IEEE Trans. on Image Processing, 20(9): pp. 2678-2683, 2011.
[12] NOVA decoder website, http://opencores.org/project,nova,overview.
[13] 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, 13(4): pp. 600-612, 2004.
[14] M. Del FabroEmail and L. Böszörmenyi, "State-of-the-art and future challenges in video scene detection: a survey," Multimedia Systems, 19(5): pp. 427-454, 2013.
[15] T.-Y. Hsieh and C.-R. Chen, "No-reference error-tolerability test methodology for image processing applications," To be presented at IEEE Int’l. Test Conf. in Asia, pp. 1-6, Aug. 2018.