Computer Generated Image Alterations Database

Matija Pul [1], Mario Vranješ [2], Zvonimir Kaprocki [3], Dejan Stefanović [3]
[1] RT-RK Institute for Information Technologies, Cara Hadrijana 10b, Osijek, Croatia
[2] University of Osijek, Faculty of Electrical Engineering, Computer Science and Information Technology, Kneza Trpimira 2B, Osijek, Croatia
[3] RT-RK Institute for Computer Based Systems, Narodnog fronta 23A, Novi Sad, Serbia
mario.vranjes@ferit.hr

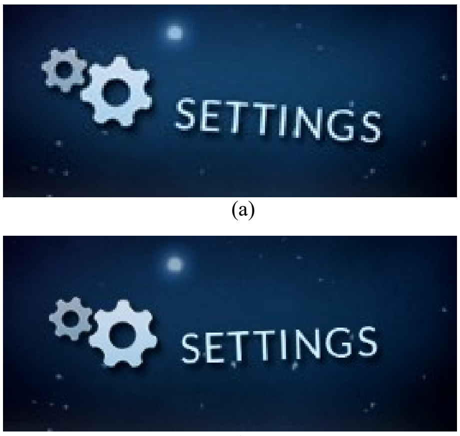
***Abstract*—Modern user interfaces (UI) use animations and special effects to appear more interesting and appealing to consumers, such is the case for modern Set-Top-Box-es (STB). In order to provide their consumers with functional and reliable products, manufacturers are required to perform functionality verification (FV) of each product. Automated FV systems are becoming a prevailing method of FV due to manual FV becoming exceedingly time consuming as the devices become more complex. Generally, full reference (FR) image comparison between the device output image and the referent image is used in order to perform FV of device under test (DUT). Common methods for image comparison proved to be ineffective when comparing images captured from UIs containing animated objects or dynamic background. In order to provide manufacturers with high-end FV systems, it is necessary to develop new methods for image comparison capable of comparing images despite alterations caused by UI animations. Development of such algorithms requires large image databases which effectively simulates such image alterations. In this paper we present new Computer Generated Image Alterations Database (CGIAD) which contains 10000 test images with alterations caused by object transformations commonly used in UI animations (rotation, scaling, occlusion, and translation). CGIAD is freely available to the research community at http://www.rt-rk.com/other/CGIAD.html.**

***Keywords*—Functionality Verification; Image Database; Image Alterations; Full Reference; Set-Top-Box**

I. INTRODUCTION

With the proliferation of TV systems, Set-Top-Box (STB) technology has advanced enough to support more complex interfaces and new ways for user interaction, such as speech recognition and animated user interfaces (UI) [1]–[4]. This has put a lot of strain on the manufacturers and created an increase in demand for more cutting-edge functionality verification (FV) systems [5], [6]. Manufacturers aim to provide their consumers with functional and reliable products. In order to achieve that, manufacturers are required to perform FV of each product before handing it over to the consumer. Manual FV requires substantial human effort and time, whereas automated systems can test multiple devices simultaneously, significantly reducing verification time [7]. Most common method used in automated FV systems for STBs is a full reference (FR) method which compares referent images, captured from a referent device, which is deemed to be functioning properly, with test images captured from a device under test (DUT) in order to test its functionality.

Dynamic UIs tend to have an ample amount of animations and special effects in order to make the UI seem more interesting and appealing to the consumer. In that case, content of captured images depends on current state of UI in the moment image is captured (see Fig. 1.). Images whose content changes over time while the image context stays unchanged are called dynamic images. Dynamic images are especially problematic when performing FV on a STB with dynamic UI. In order to verify the UI is functioning properly, it is required to verify every UI element is properly displayed. Problem appears when trying to verify animated objects which can move around screen or even change shape and orientation. FV systems based on FR image comparison require advanced algorithms capable of comparing dynamic images. Testing of such advanced algorithms requires a referent device and a DUT to be available for capturing images which would then serve for FV. In order to avoid requiring such a setup during development and optimization of said algorithms it is possible to create a large database of images simulating alterations often present in modern STBs dynamic UIs. That is why we present a new database called Computer Generated Image Alterations Database (CGIAD), which contains 10000 images and is freely available to the research community at http://www.rt-rk.com/other/CGIAD.html. Every image in CGIAD is experiencing alterations caused by transformations which can be often found in dynamic UI animations, such as rotation, scaling, occlusion, and translation. To the best of our knowledge there are no existing freely publicly available databases with focus on dynamic image comparison.



(a)

(b)

Figure 1. An example of a single animated object captured at two different moments: (a) moment 1, (b) moment 2. The object is experiencing the following transformations: rotation, scaling, translation and occlusion.

The rest of this paper is organized as follows. Section II gives more information about CGIAD and how it was created. Section III gives results of standard image quality evaluation/image similarity comparison algorithms when applied to CGIAD. Conclusions are given in Section IV.

II. Computer Generated Image Alterations Database

CGIAD was created in order to provide materials for development and testing of dynamic image comparison algorithms. Images provided in this database are aimed to simulate image alterations caused by animations often observed on dynamic UIs.

A. *Image alteration tool*

In order to create CGIAD, a special tool was created, capable of applying transformations to any given object. The tool utilizes OpenCV (version 3.4.1) library [8] to work with images. Dynamic UIs can contain any type of animation, making it impossible to simulate every possible object behavior. In order to capture as many possible behaviors, the created tool implements four basic transformations: rotation, scaling, occlusion, and translation. Transformations can be applied both individually and combined, with a provided transformation intensity range. It is possible to set the number of images generated for each test case scenario as well as intensity range for each transformation. This allows for tool automatization which was required in order to create a large image database.

B. *CGIAD composition*

Transformations mentioned in the previous sub-section are always applied on a pair of images. Image pair consists of referent object and false object. Referent object is used in the referent image and true positive (TP) test cases while the false object is used in true negative (TN) test cases. Examples of object pairs can be seen in Fig. 2. For each object pair, five test case scenarios are generated: rotation only, scaling only, occlusion only, translation only, and combined transformations.

Each transformation alters the image, making it less recognizable. Rotation can cause drastic changes to the object, depending more on object shape rather than rotation intensity, as demonstrated in Fig. 3(b). Round objects tend to experience far

(a)

(b)

(c)

(d)

(e)

Figure 3. Example of image alterations caused to the referent object by the following transformations: (a) no transformations (b) 30° rotation, (c) 20% downscaling, (d) 10% particle occlusion, and (e) combination of 14° rotation, 12% downscaling, and 8% occlusion.

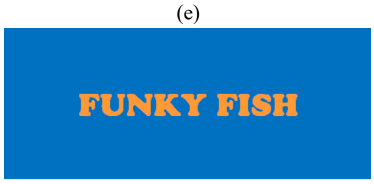
(a)

(c)

(e)

(b)

(d)



(f)

Figure 2. Example of object pairs used in true positive and true negative tests: (a) referent object in first pair, (b) false object in first pair, (c) referent object in second pair, (d) false object in second pair, (e) referent object in third pair, and (f) false object in third pair.

less alterations than objects with straight edges. Scaling will always cause significant alterations, proportional to scaling intensity, as can be seen in Fig. 3(c). Object shape plays a less influential role in scaling than in rotation. Most important factor in object scaling is the scaling algorithm which determines how precise the approximated data is and how much information is lost. It is safe to assume scaling is the most problematic transformation from the perspective of dynamic image comparison. Occlusion is a less common transformation but potentially the most dangerous one, because it causes complete information loss on the area it appears on. In order to simulate object occlusion, the tool randomly applies small particles, 5 × 5 pixels in size, to the image until a required percentage of image is covered in particles. Occlusion example is shown in Fig. 3(d). Translation is the simplest transformation and does not cause any alterations to the object itself but is used in most animations. Still object translation presents a significant problem for standard image comparison algorithms based on pixel-by-pixel comparison.

When performing FV, the manufacturers must be 100% sure the device is fully functional before handing it over to the customer. This means the precision must be 100% when testing devices, i.e. number of false positives must be zero. For this reason we created object pairs which are used for TP and TN tests. Objects used in pairs are evidently different to a human eye but are also fairly similar in content and context (see Fig. 2.). Goal was to make it challenging for algorithms to find differences between objects while also keeping them different enough that a human could easily distinguish them. TP and TN tests are required in order to determine algorithm accuracy and precision.

CGIAD consists of 10000 test images with alterations caused by applying transformations to 10 different object pairs. Images are 1920×1080 pixels in size and contain the object experiencing transformations along with the text describing transformation intensities and image name (see Fig. 4.). There are a 1000 test images for each object pair, where 500 images serve for TP tests and the other 500 serve for TN tests. Aim of TP and TN tests is to differ between objects which means there must be no other factors impacting the comparison other than changes in objects themselves. Because of this, the same transformations are applied on both TP and TN objects. This way we ensure the objects are experiencing same behavior when performing comparison with the referent object, i.e. any change

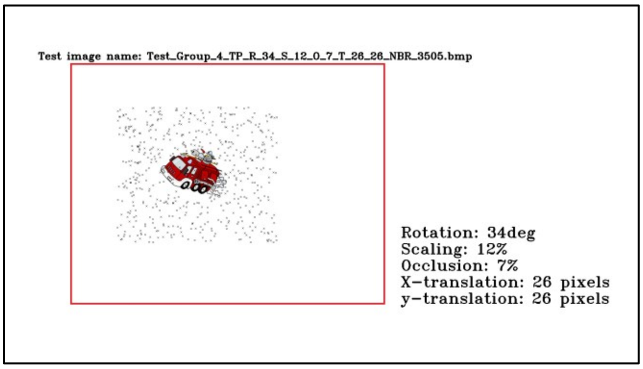


Figure 4. Example of an image from CGIAD. Red square was added in order to portray region-of-interest (ROI) used for testing purposes.

caused by transformations will appear in both images with the same degree of intensity. Test case scenario format used for CGIAD is: 50 images with rotation only, 50 images with scaling only, 50 images with occlusion only, 50 images with translation only, and 300 images with combined transformations. This format was chosen because it allows the user to test algorithm behavior for each transformation type both individually and combined. Individual transformation cases can be especially useful during algorithm development whereas combined transformation are more probable to appear in real case scenarios and can be used for algorithm assessment.

When applying transformation to the object, the tool picks a unique random transformation intensity value from a given range in order to avoid duplicate images. This applies to all images not containing occlusion, since particle locations are also randomized so the images with the same occlusion intensity values have different image alterations. All transformations can have a positive or negative intensity value except occlusion, which can only be positive. Maximum intensity values used for creation of CGIAD are: rotation up to 40°, scaling up to 20%, occlusion up to 10%, and translation up to 30 pixels horizontally and vertically.

III. Performance Measurements Of Existing Still Image Similarity Algorithms On CGIAD

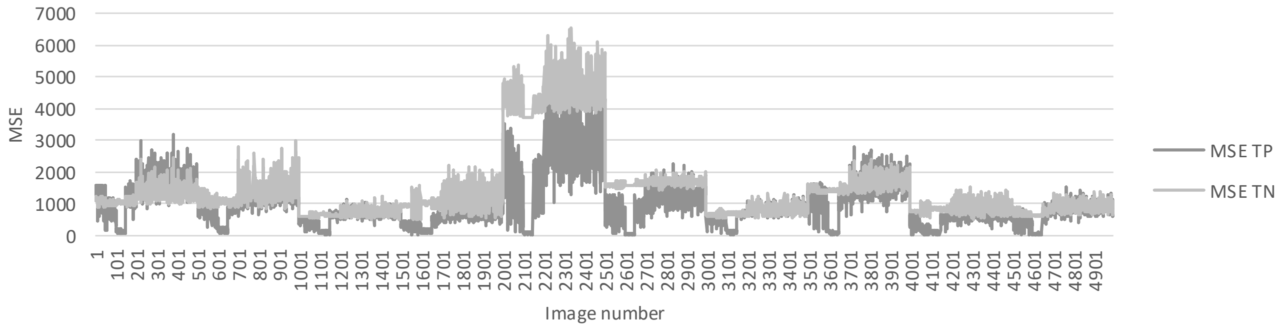
In our experiments we wanted to examine whether the conventional FR image similarity metrics (i.e. objective image quality metrics when original and altered images are compared) can be applied for comparison of dynamic images like those from CGIAD, all with the purpose of applying them in automated systems for FV of STBs with dynamic UIs.

FR image comparison is often performed using objective image quality metrics. In our experiments we have tested three different image quality metrics, Mean Square Error (MSE), Peak Signal-to-Noise Ratio (PSNR) [9], and Structural Similarity Index (SSIM) [10]. Duo to its simplicity, MSE is commonly used FR objective image quality measure and is computed by averaging the squared intensity of the difference between the referent image and corresponding test image pixels. Similar to MSE, PSNR estimates the difference between images based on the pixel difference. The MSE represents the cumulative squared error between the compressed and the original image, whereas PSNR represents a measure of the peak error. Unlike MSE and PSNR, SSIM does not observe each pixel individually and instead observes groups of pixels in order to get a better representation of the image making it less sensitive to image alterations. When used for image quality estimation, SSIM metric uses structural distortion in video as an estimate of the perceived visual distortion. SSIM shows high performance when tested on images/videos from different databases. The SSIM score increases as the similarity of the referent and the test image rises and it is 1 for the completely same images.

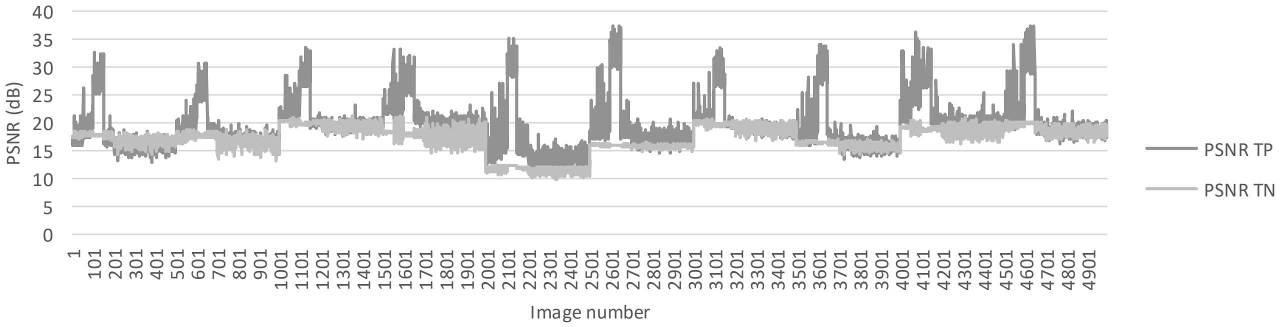
Fig. 5. shows image comparison results for MSE, PSNR, and SSIM algorithms when used on CGIAD. As can be observed from the graphs, none of the three commonly used image comparison methods can efficiently resemble between TP and TN test images. MSE value grows proportionally with image difference, meaning the closer the value is to 0 the greater the

image similarity. Analyzing the results for MSE shows it scored best results for TP images in cases where only occlusion was applied. Reason why this happens is because, during testing, all images are using same region-of-interest (ROI), size of 950×725 pixels with a starting point coordinates (200, 175), which is larger than the object itself in order to avoid object clipping (see Fig. 4.). Applying transformations to object can cause it to change size and location so the ROI has to be large enough to contain all possible object locations after transformations. Only the pixels within ROI are observed during image comparison. Since MSE is a pixel-by-pixel comparison method, it measures the amount of different pixels and the only pixels changing during occlusion are the ones where the particles have landed. They might cover up to 10% of the object surface but when taking into account all the unaltered pixels from the entire ROI it makes the alterations caused by occlusion seem a lot less significant. Similar to MSE, PSNR scored somewhat better results for TP images. Automated FV systems require high precision algorithms, which means there must be a single application-specific threshold which could be used to make a binary decision based on the image comparison results. Observing the results from Fig. 5(b), it is clear that no such threshold exists. This means the PSNR is not suitable for use in FV of devices with dynamic images. SSIM results shown in Fig. 5(c) also show overlapping between TP and TN images. Even though SSIM is a more advanced method for image comparison, which is not based solely on pixel-by-pixel comparison, it fails to recognize the same object when animated. On the other hand, it is reasonable, since the main purpose of the SSIM is image perceptual quality evaluation, not exactly similarity measurement for dynamic images.

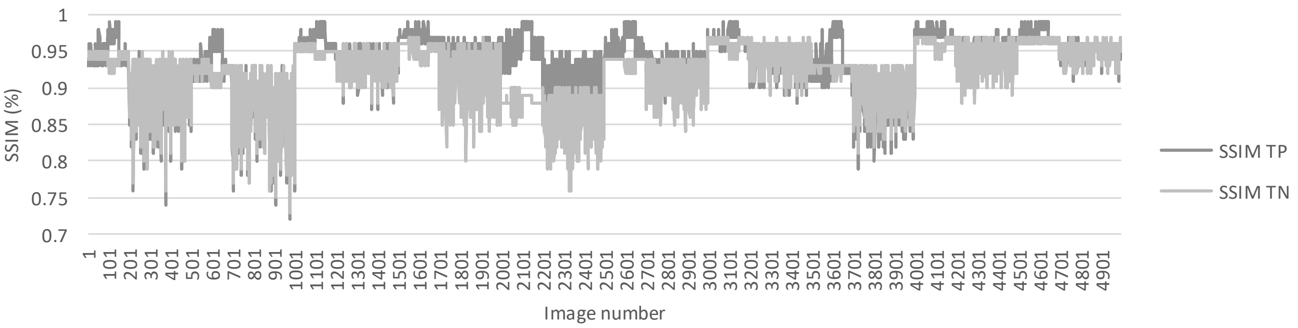
None of the tested methods have proved capable of comparing dynamic images making them unreliable for FV of devices with dynamic UI. Algorithm reliability is based not only on the ability to recognize if the desired object has appeared but



(a)



(b)



(c)

Figure 5. Image comparison results on CGIAD for (a) MSE, (b) PSNR, and (c) SSIM [10] algorithms.

also on the ability of recognize when a similar but different object has appeared. Ideal algorithm would have a significant difference between results for TP and TN test images proving it can recognize that a similar object has appeared while also realizing that it is not the correct object. Such algorithm would be both precise and accurate, making it suitable for use in automated FV systems for devices with dynamic UI. Also, such algorithm could be developed with the help of CGIAD.

IV. Conclusion

Most commonly used methods for image comparison proved tremendously inaccurate when comparing dynamic images. In order to provide manufacturers with high-end FV systems, new, more precise and more reliable methods need to be developed. CGIAD contains 10000 test images with image alterations generated by simulating the most common types of dynamic UI animations (rotation, scaling, occlusion, and translation). Database contains TP and TN test cases for 10 unique object pairs. CGIAD aims to provide developers with material suitable for research and development of new methods for dynamic image comparison. Presented database is freely available at http://www.rt-rk.com/other/CGIAD.html. For our future work we plan to develop one such method using CGIAD.

Acknowledgment

This work was partially supported by J.J. Strossmayer University of Osijek business fund through the internal competition for the research and artistic projects „IZIP-2016“ and partially by the Ministry of Education, Science and Technological Development of the Republic of Serbia, under grant number: III_044009_6.

References

[1] S. Pravin and R. BalaKrishnan, “Set top box system with android support using Embedded Linux operating systempaper,” in *IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012)*, 2012, pp. 474–478.

[2] K. Fujita, H. Kuwano, T. Tsuzuki, Y. Ono, and T. Ishihara, “A new digital TV interface employing speech recognition,” *IEEE Trans. Consum. Electron.*, vol. 49, no. 3, pp. 765–769, Aug. 2003.

[3] D. Grbić, M. Vranješ, B. Kovačević, and M. Milošević, “Hybrid Electronic Program Guide Application for Digital TV Receiver,” presented at the 7th IEEE International Conference of Consumer Electronics, 2017.

[4] N. Jovanov, D. Kovacevic, S. Pejic, and G. Velikic, “One solution of visualizing Internet based EPG data combined with additional Internet content on an Android based set-top box,” in *2014 IEEE Fourth International Conference on Consumer Electronics Berlin (ICCE-Berlin)*, 2014, pp. 114–116.

[5] N. Nikić, M. Herceg, V. Peković, and N. Šoškić, “System for DASH support verification in HbbTV environment,” presented at the 7th IEEE International Conference of Consumer Electronics, 2017.

[6] M. Plenković, M. Herceg, V. Peković, and S. Novak, “Coupling of HbbTV test system with automatic testing subsystem,” presented at the 7th IEEE International Conference of Consumer Electronics, 2017.

[7] D. Marijan, N. Teslic, M. Temerinac, and V. Pekovic, “On the effectiveness of the system validation based on the black box testing methodology,” in *Testing and Diagnosis, 2009. ICTD 2009. IEEE Circuits and Systems International Conference on*, 2009, pp. 1–4.

[8] “OpenCV library.” [Online]. Available: https://opencv.org/releases.html. [Accessed: 07-May-2018].

[9] J.-B. Martens and L. Meesters, “Image dissimilarity,” *Signal Process.*, vol. 70, no. 3, pp. 155–176, 1998.

[10] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.