Image Features Correlation with the Impression Curve for Automatic Evaluation of the Computer Game Level Design

Jarosław Andrzejczak[1,2][0000−0002−4124−0110], Olgierd Jaros[1], Rafał Szrajber[1,3][0000−0003−2777−0251], and Adam Wojciechowski[1][0000−0003−3786−7225]

[1] Institute of Information Technology, Lodz University of Technology, 215 Wólczańska Street, 90-924 Lodz, Poland
http://it.p.lodz.pl
[2] jaroslaw.andrzejczak@p.lodz.pl
[3] rafal.szrajber@p.lodz.pl

Abstract. In this study, we present the confirmation of existence of the correlation of the image features with the computer game level Impression Curve. Even a single image feature can describe the impression value with good precision (significant strong relationship, Pearson $r > 0{,}5$). Best results were obtained using by combining several image features using multiple regression (significant very strong positive relationship, Pearson $r = 0{,}82$ at best). We also analyze the different set of image features at different level design stages (from blockout to final design) where significant correlation (strong to very strong) was observed regardless of the level design variant. Thanks to the study results, the user impression of virtual 3D space, can be estimated with a high degree of certainty by automatic evaluation using image analysis.

Keywords: image analysis · Virtual Reality · Impression Curve · level design · automatic evaluation

1 Introduction

In [1] study, we have shown that Virtual Reality space affects different users in a similar way. That sense can be stored and described as Impression Curve[4] for this space. Therefore, Impression Curve can be used in 3D VR space evaluation such as 3D level design. Still, it requires tests with many users to gather proper data. It would be a great improvement if designers could estimate the sense of 3D space during the development process and then verify it at the end with the users. Especially with the growing popularity of level designs generated by algorithms [16]. We focused our efforts to provide such computationally low cost tool for 3D space evaluation in the context of estimating user experience.

[4] Impression Curve is a measure of the visual diversity and attractiveness of a game level. It assesses subjective attraction of a given space. For the detailed information about the Impression Curve, its acquisition method, its strengths and weaknesses in the domain of the 3D space evaluation, please refer to [1].

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_49

The purpose of the research was to verify the existence of the correlation of the image features (gathered using automatic image analysis) with the Impression Curve (obtained during previous studies conducted on 112 people). The study described in this article involves examining the impact of various image features such as mean brightness and contrast, features based on saliency and movement maps (such as complexity or density), as well as descriptive statistics like entropy, skewness and kurtosis.

The contributions to research concerning automatic evaluation of the immersive Virtual Reality space, especially in case of the Impression Curve estimation presented in this article, are:

- Confirmation of the existence of a correlation between data gathered using image analysis and user-generated Impression Curve.
- Tests verifying the correlation between individual image feature and the Impression Curve for the VR space.
- Tests verifying the correlation between combined image features and the Impression Curve for the VR space.
- Analysis of usability of each image features depending on the level design stages and changing factors of the 3D space.
- Proposition of the best image features (with the highest correlation values with Impression Curve) for evaluation of individual level design stages.

We start with a related work overview in the domain of image analysis for feature extraction in the next section. Then we describe hypotheses and an evaluation method. Next, both test results and their discussion will be presented, as well as observations about data gathered. Finally, ideas for further development and final conclusions will be given.

2 Image features

There are many image features available to consider in terms of image analysis for automatic feature extraction and image description. Our goal was to test as diverse set of features as possible. The three groups of features were used: color and luminance-based (such as mean brightness, mean color contrast) [6], features based on saliency and motion maps (such as balance and density) [5], as well as descriptive statistics (entropy, skewness and kurtosis) [7]. Therefore, a total of thirteen features were selected for this study:

- **Color and luminance group:** Average Contrast, Average Luminance and Average Saturation.
- **Saliency and motion maps group:** Alignment Complexity, Balance Complexity, Density Complexity, Grouping Complexity, Size Complexity and Total Complexity.
- **Descriptive statistics group:** Entropy, Kurtosis, Skewness and Fractal Complexity.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_49

For the calculations of the image **Average Contrast**, **Average Luminance** and **Average Saturation**, the definitions for the HSL color palette were used. Average Contrast was calculated using the mean square of the Luminance of individual pixels [6].

Image features from the second group are based on classification and analysis of areas indicated in saliency maps [2] and motion maps [8]. Those maps are combined (with a weight of 50% of each, as we considered them equally important) and classified to be used with the metrics described in [5]. This stage requires the greatest number of computations. We start with the creation of the saliency map using the fast background detection algorithm [2], which is then denoised using the method described in [3]. The result is a black and white image, with white pixels representing the relevant ones. A motion map is created as a difference of the pixels of two subsequent video frames converted to grayscale with a Gaussian blur applied to them (which allows limiting the influence of details and noise on motion detection) [4]. The resulting image is denoised and thresholded to obtain a black and white image and combined with the saliency map to obtain the final visual attention saliency map [14]. Then the classification of regions, objects and their contours as well as shape recognition is made.

Regions of attention (representing grouped objects) and their centroids are calculated using K-Means with 30 starting points (pixels) picked randomly on visual attention saliency map white pixels. For each iteration, the closest region centroid for each point is calculated and the region centroid weights are updated. The algorithm runs for 1000 epochs or until each region centroid remains unchanged in two subsequent epochs. During this process, centroids, which for two ages were not the closest one for any point, are permanently removed from the set to optimize the calculations. Centroids calculated for one frame become the starting points for the next frame, with one new random starting point added (to allow the new area recognition).

Objects of attention are found by applying erosion filter and OpenCV shape detection [10] on the final visual attention saliency map. Next, the object's contour is calculated using the contour approximation method [10]. For each of the identified object, a centroid is calculated. Please note that object's centroid is usually different from region centroid, as one region can contain many objects.

Localized object's contours are therefore used for shape recognition [10]. Only simple geometric shapes are taken into account, and every object with a number of vertices greater than or equal to five is classified as a circle (for the purpose of further analysis).

All of the above final visual attention saliency map characteristic is then used with the metrics for UI complexity analysis described in [5]. Each metric gives a final score in the range [0,1] where a score closer to zero means less complexity. The **Alignment Complexity** determines the complexity of the interface in terms of the position of the found shapes relative to each other. The evaluation consists of the calculation of the local and global alignment coefficients for grouped and ungrouped objects. The **Density Complexity** determines the comparison of the visual attention object size to the entire image frame size.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_49

The **Balance Complexity** describes the distribution of visual attention objects on the quarters of the screen. It is calculated as the arithmetic mean of two mean values: the proportion of the number of objects between pairs of quarters and the proportion of the size of objects between pairs of quarters. The **Size Complexity** is calculated due to the grouping of objects on the screen in terms of shape. For each shape type, the number of occurrences of the size of objects is checked. Then The sum of the occurrences of unique object regions is divided by the number of objects in the particular group of shapes. The **Grouping Complexity** determines how many of the objects are grouped into shape type groups. It is the sum of the ratio of ungrouped objects to all occurring and the number of groups of shapes occurring in the region of objects from all possible shapes types. The **Total Complexity** is a combined metric of all previous with weights as proposed in [5]:

$$TotalComplexity = \begin{array}{l} 0,84 \times Alignment + 0,76 \times Balance + \\ 0,8 \times Density + 0,72 \times Size + 0,88 \times Grouping \end{array} \tag{1}$$

The third group of image features is based on statistical descriptors of a data set's distribution. The **Skewness** is a measure of the asymmetry of a distribution of the mean. The higher the Skewness, the more asymmetric data distribution. The **Kurtosis** is a measure of how results are concentrated around the mean. The high Kurtosis value would suggest outliers in the data set and low Kurtosis value the lack of outliers [11]. The **Entropy** of an image is used as a measure of the amount of information it contains [7]. The more detailed the image, the higher the value of the Entropy will be. Entropy, Kurtosis and Skewness were counted separately for Hue, Saturation and Luminosity as they operate on the single variable (grayscale image as input). The **Fractal Complexity** is a measure of self-similarity. It determines how much it is possible to break an image or fractal into parts that are (approximately) a reduced copy of the whole. This parameter was used to assess the complexity of the image [9] [5].

3 Evaluation

The goal of the evaluation was to verify the existence of the correlation of the image features (gathered using automatic image analysis) with the Impression Curve. For this purpose, the Pearson and Spearman correlation were used [13]. All the level design stages as well as the influential factors on the 3D space impression (such as lightening condition changes, geometrical and material changes) described in [1] were used (Fig. 1).

[5] At this stage of the Impression Curve automatic evaluation study, we have used the controlled 3D space designs to minimize the influence of the such factors as action, gameplay rules and restrictions, story and lore present in commercial game designs. After confirmation of existence of the correlation of the image features with the computer game level Impression Curve described in this article, we moved to testing level design from popular games. The results of this study will be published in the future, as it is in development at the time of writing this article.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_49

The study was divided into two parts. First, the correlation of the individual image features with the Impression Curve was analyzed. After that, image features with the highest correlation value were combined into sets and once again tested for correlation with Impression Curve to see if there is any gain in the strength of the correlation.

The hypotheses in individual parts were as follows:

1. First part: there is a significant correlation (positive or negative) between an individual image feature and the Impression Curve for the same VR space.
2. Second part: the correlation (positive or negative) with the Impression Curve is higher for the combined image features than for the individual image features.
3. Additional observation: different set of image features presents the highest correlation values for different level design stages.

What is more, different level design stages and changing factors of the 3D space (for example: lightening condition, geometrical detail or material changes) of the same game level allow us to observe if there is any difference in correlation between data gathered using image analysis and user-generated Impression Curve. Thanks to this, we were able to point out the best automatic evaluation measures in the form of selected image features, to use at each design stage (blockout, models without materials, textured models as well as lightning and atmospheric effects such as rain).

The twelve level variants showing successive design stages were used according to our previous research, described in details in [1]. There were as follows: simple blockout (A), advanced blockout (B), main models without materials (C), main models with monochromatic materials (D) and final materials (E) as well as with extra fine detailed models (called final level version) (F), main models with geometrical changes (G) and final level with changes of visual factors as lightening condition (L), weather condition (W), different materials (M), added expression (X) as well as with extra models and objects in the environment (O). Existence of correlation between image features and Impression Curve values would allow creation of a tool to automatically estimate Impression Curve for a VR space with a high degree of probability. And as a result, to automatically evaluate expected user impression even on an early Virtual Reality space design stage.

4 Results and analysis

During the study, hundreds of correlation plots were gathered and analyzed. We assumed that per frame comparison will be sensitive to rapid image changes, effecting low or no correlation at all. That is why, the mean and median of an image feature for a few consecutive frames were calculated. A small range of 4-5 frames allow us to eliminate minor fluctuations, where a larger range of 20-30 frames softened the charts quite significantly. However, a larger range considerably reduces the number of data samples, which had an impact on the significance value

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_49

Fig. 1. The twelve level variants showing successive design stages used in this study for image analysis and correlation with Impression Curve. A - simple blockout; B - advanced blockout; C - models without materials; D - models with monochromatic materials; E - models with final materials; F - final level version; G - geometrical changes; L - lightening condition changes; W - weather changes; M - material changes; X - expression added; O - extra models added.

p. Thus, we started from a range of four frames and increased this interval by four from that point. As a result, four to twenty frames, we observed increased correlation value for most of the image features while preserving low value of $p <$ 0,05. For frame range greater than twenty, results were not significant anymore ($p >$ 0,05). Also, above this point, the correlation value for many image features dropped below the value of 0,3. Thus, we choose a range of twenty frames for our study, as it shows the highest correlation values with significance $p <$ 0,05 (in many cases $p <$ 0,01). In the other hand, we gathered image data more often (thirty times per second - video recorded with a 30 FPS frame rate) than during study with users. Thus, the Impression Curve data had to be interpolated between measure points (as we assumed linear change). This way we were able to compare this data even per frame.

The experiment stages were as follows: first, for each of the video game level variants the Impression Curve data (gathered with users) was interpolated between the measure points to match the frequency of data calculated using image analysis for this level variant walkthrough video; next, the image features were calculated and refined using respectively mean and median for 20 subsequent frame intervals; finally, the Pearson and Spearman correlation between those data were calculated.

The recordings of twelve variants of the video game level variants (used in [1]), including twenty-nine thousand three hundred and thirty-nine frames in total, were analyzed. As a result, thirty-six data sets were obtained and used to generate two hundred and ninety-nine correlation plots.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_49

4.1 Individual image features correlation

The first part of the study involved testing each of the thirteen image features individually for correlation with an interpolated Impression Curve for each of twelve variants of the video game level described earlier. The result of a single feature-variant pair was stored in numerical way and also as a correlation plot for easier analysis (Fig. 2). Each data point in the graph shows respectively the mean or median (depending on which one was used) over an interval of 20 frames of the video data. The feature values are marked in red, while the values of the Impression Curve are marked in green. The charts contain the calculated Pearson correlation for a whole Impression Curve. When this value is below 0,5 the Spearman correlation is calculated as well to compensate possible outliers and check for nonlinear relation. Two numbers are presented for each correlation. The first is the mean correlation value, the second is the calculated p value of this correlation.

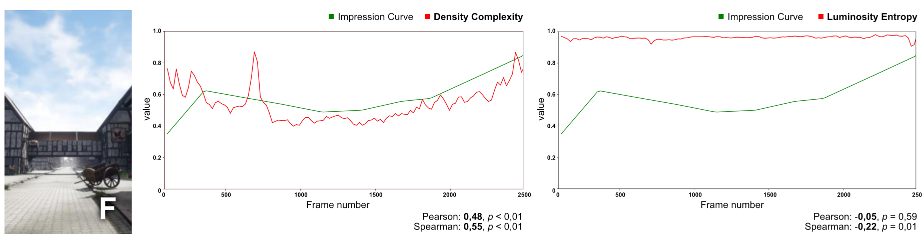

Fig. 2. Correlation plot examples for final level design variant (F variant, on the left). Two image feature correlation plots are presented: one with significant strong positive relationship - Density Complexity (Pearson r = 0,48 with $p < 0{,}01$, center) other with no significant linear relationship and weak non-linear relationship - Luminosity Entropy (Pearson r = -0,05 with $p = 0{,}59$, right). A linear relationship can be observed for Density Complexity.

Then the correlation values of every image feature tested for a single level design variant were juxtaposed with each other (Table 1 shows the results for only one variant as an example - the same was done for each of twelve level design variants).

We observed many significant correlation values (positive and negative) between image features and Impression Curve value. Observation varied from a few weak relationships (r value between 0,20 and 0,29) to moderate relationship in most cases (r value between 0,30 and 0,39) and even over a dozen strong relationship (r value between 0,40 and 0,69). There was not a single variant without at least one significantly related image feature, and in most cases there were several moderate relationships. What is more, some image features tend to correlate more often than others, where others given at least weak relationship only once or twice (Table 2). We did not observe a significant difference between

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_49

Table 1. Pearson's correlation values for individual image features of the final version of the level (F). Feature values were calculated respectively as the mean and median for the intervals of twenty frames. The highest correlation results are marked with a gray background color and bold text. The significant p values are marked with a gray background color. We can observe that the same image features show the highest correlation and similar values for both the mean and the median, with only one feature (Grouping Complexity) presenting lower correlation using the median. r - Pearson correlation coefficient value; p - significance value.

	Mean		Median	
Image feature	r	p	r	p
Alignment Complexity	0,07	0,463	0,06	0,516
Balance Complexity	**0,33**	<0,001	**0,32**	p <0,001
Density Complexity	**0,48**	<0,001	**0,52**	p <0,001
Grouping Complexity	**0,28**	0,002	0,16	0,087
Size Complexity	**0,43**	<0,001	**0,46**	p <0,001
Total Complexity	**0,53**	<0,001	**0,48**	p <0,001
Average Contrast	**-0,28**	0,002	**-0,28**	0,002
Average Luminance	0,01	0,951	-0,01	0,908
Average Saturation	**-0,36**	<0,001	**-0,35**	0,000
Fractal Complexity	-0,05	0,551	-0,06	0,544
Hue Entropy	**0,35**	<0,001	**0,35**	p <0,001
Hue Kurtosis	-0,04	0,651	-0,03	0,737
Hue Skewness	0,05	0,582	0,07	0,457
Saturation Entropy	-0,16	0,083	-0,15	0,103
Saturation Kurtosis	-0,03	0,719	-0,07	0,477
Saturation Skewness	-0,04	0,701	-0,04	0,652
Luminosity Entropy	0,03	0,719	0,05	0,587
Luminosity Kurtosis	**0,33**	<0,001	**0,33**	p <0,001
Luminosity Skewness	0,06	0,508	0,07	0,456

mean and median values ($t-test\ p = 0{,}52$) thus only median will be used in further analysis as less valuable for outliers.

For all but one image features, we observe no significant difference between Pearson and Spearman correlation coefficient values, which suggest a linear nature of the relationship. Thus, in further combined image features we focused on Pearson correlation coefficient as linear relationship is more desired for the future video game level design automatic evaluation system. Only for Density Complexity feature, we observed significant difference ($t-test\ p = 0{,}05$) between Pearson and Spearman results with Spearman correlation coefficient values being higher most of the time giving moderate to high positive relationship (also with much lower p value). This indicates the existence of a non-linear relationship between Density Complexity feature and the Impression Curve.

The results are dominated by a positive correlation, with six image features tending to present a negative relationship more often than positive. Those are: Grouping Complexity, Fractal Complexity, Average Contrast, Average Saturation and Entropy (for Saturation and Luminosity). Most of them present many moderate to strong relationships (also variants with low correlation value results were not significant with $p > 0{,}05$). The highest single image features correla-

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_49

Table 2. Pearson's correlation values for individual image features for all twelve level design variants. Feature values were calculated as median for the intervals of twenty frames. The significant correlation results (with $p =< 0{,}01$) are marked with a grayscale background color (the darker the color, the higher the correlation value) and bold text. Strong relationship (r value between 0,40 and 0,69) was outlined with a white text color. We can observe that some image features as Size Complexity or Grouping Complexity tend to present high correlation value in many variants. A - simple blockout; B - advanced blockout; C - models without materials; D - models with monochromatic materials; E - models with final materials; F - final level version; G - geometrical changes; L - lightening condition changes; W - weather changes; M - material changes; X - expression added; O - extra models added.

	Level Design Variant											
Image Feature	A	B	C	D	E	F	G	L	W	M	X	O
Alignment Complexity	-0,15	-0,01	-0,04	0,13	0,14	0,06	0,10	0,09	**-0,25**	0,02	-0,07	-0,10
Balance Complexity	0,04	0,06	**0,33**	**0,25**	0,18	**0,32**	0,08	0,14	0,19	**0,47**	**0,33**	**0,31**
Density Complexity	0,23	0,17	0,20	**0,33**	**0,33**	**0,52**	0,18	0,20	**0,39**	0,20	-0,09	0,15
Grouping Complexity	-0,14	**-0,25**	**-0,40**	**-0,33**	0,06	0,16	**-0,27**	**-0,36**	**-0,24**	-0,02	0,18	0,06
Size Complexity	-0,21	-0,19	**-0,36**	**-0,32**	**0,57**	**0,46**	-0,16	**-0,42**	**0,38**	**0,24**	**0,34**	**0,39**
Total Complexity	-0,21	0,05	0,00	0,23	**0,38**	**0,48**	-0,06	-0,20	0,19	**0,28**	0,04	0,14
Average Contrast	**-0,55**	-0,28	-0,14	0,10	-0,21	-0,28	-0,30	-0,37	-0,27	-0,06	0,12	0,02
Average Luminance	-0,10	-0,15	0,42	-0,24	0,25	-0,01	0,03	-0,03	-0,03	0,19	0,03	0,08
Average Saturation	-0,41	-0,10	-0,20	0,47	-0,34	-0,35	-0,40	0,26	0,29	-0,31	-0,07	-0,16
Fractal Complexity	0,22	0,01	-0,26	-0,25	-0,02	-0,06	0,00	-0,06	-0,28	-0,05	-0,21	-0,19
Hue Entropy	-0,08	-0,18	0,11	0,32	0,27	0,35	-0,12	0,21	0,06	0,30	0,32	0,30
Hue Kurtosis	0,28	0,29	0,01	0,37	0,37	-0,03	0,17	0,06	0,43	0,05	-0,11	-0,04
Hue Skewness	-0,25	-0,32	0,03	0,47	0,38	0,07	-0,04	-0,08	-0,56	0,12	-0,05	-0,09
Saturation Entropy	-0,31	-0,07	-0,14	0,05	-0,38	-0,15	-0,25	0,19	0,13	-0,22	0,21	0,00
Saturation Kurtosis	0,04	0,06	0,23	-0,55	0,21	-0,07	0,17	-0,05	0,28	0,34	-0,09	0,07
Saturation Skewness	0,06	-0,01	0,29	-0,49	0,20	-0,04	0,22	0,29	0,35	0,37	0,11	0,21
Luminosity Entropy	-0,36	-0,32	0,07	0,16	0,05	0,05	-0,25	-0,22	-0,43	0,16	0,07	0,16
Luminosity Kurtosis	0,27	0,18	0,35	0,12	0,55	0,33	0,15	0,53	0,22	0,15	0,12	0,09
Luminosity Skewness	0,17	0,13	-0,41	0,13	-0,33	0,07	0,00	0,45	0,10	-0,12	-0,09	-0,11

tion value observed was 0,57 (strong positive relationship, $p < 0{,}01$) for a Size Complexity feature in variant of models with the final materials (E).

We also observed that the earlier the level creation stage, the lower the correlation values of most image features (Table 2). The materials used in the virtual space design has a great influence on the correlation value. In the case of variant C (3D models without materials), a significant strong relationship weak relationship with the Average Luminance can be noticed. This correlation decreases after adding materials to the models (variants D with monochromatic materials and E with final materials) effecting with no significant relation in final level variant (F with lightning). Similar observation can be made with Saturation Kurtosis and Saturation Skewness giving the highest correlation values for variant with monochromatic materials (D) and also no significant relation in the final level variant. Another interesting observation can be made in first design stage (simple blockout - variant A). In such a simple block design, the color-based image

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_49

features gave the highest correlation values with significant strong negative relationship for Average Contrast (Pearson r = -0,55, p < 0,01). This relation weakens with the addition of final models and textures. It is also worth paying attention to the fact that with the appearance of the final materials, the sign of the correlation for the Size Complexity image feature changes from negative to positive relationship.

It must be remembered that the value at a given point for the correlating images feature shows the general tendency of the Impression Curve (increase or decrease of it) - not the exact values of it. To reproduce the value of the curve, it is necessary to know its value at one point at least. At the same time, the change in perception of virtual space (increase or decrease) is a feature shared by users (as shown in the research presented in [1]), while the assignment of a numerical value to the Impression Curve may depend on the user and the definition of the rating scale. Therefore, the use of the Immersion Curve value change in the automatic evaluation system of the game level is not only a more reliable, but also more universal (less dependent on the user).

4.2 Combined image features correlation

Among the image features tested, the most common correlation between them and Impression Curve can be observed in seven cases (Table 2). They were divided into two groups:

- The most promising that gives the highest correlation values, especially in final level design variant (F). Those are: **Density Complexity, Size Complexity, Total Complexity and Balance Complexity**. This group formed a base set for all the combined set (and will be referred to as DTSBC hereinafter).
- The second most promising with a little lower correlation value than the first group or high relationship with variants other than final level design (F). Those are: **Grouping Complexity, Average Contrast, Average Saturation**. They were added, in every possible combination, to the first group and checked for improvement in relationship strength.

In addition to the above, color-based image features of Entropy, Kurtosis and Skewness for Hue, Saturation and Luminosity were also included in described sets as they presented significant correlation values in different stages of design (especially in early stages A to E). Image features in those sets were combined using multiple regression. From all the combined sets, those with the best Pearson's correlation values were selected (Table 3).

There was significant strong or very strong positive relationship in all cases. The best results overall were achieved for the sets *DSTBC + Average Contrast + Average Saturation + Hue Entropy* and *DTSBC + Average Contrast + Average Saturation + Hue Entropy + Luminosity Kurtosis* where the latter works for a larger number of variants (thus it is more universal). In almost all cases, the combined feature sets correlated significantly better than the single ones

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_49

Table 3. The best Pearson's correlation values for combined image features for all twelve level design variants. Features were combined using multiple regression. Three best correlated image features: Density Complexity + Size Complexity + Total Complexity + Balance Complexity (called DTSBC for short) were bases for all four combined sets. The correlation results are color-coded as a heatmap with a grayscale background color (the darker the color, the higher the correlation value). Very strong relationship (r value higher than 0,70) was outlined with a white text color and bold text. All the results were significant ($p =< 0{,}01$). There was significant correlation in all cases, where the best results were achieved for sets DTSBC + Average Contrast + Average Saturation + Hue Entropy + Luminosity Kurtosis. We can observe that the more advanced level design stage (B to F) the stronger the correlation. Also, the combination of HSL Entropy, Kurtosis and Skewness can be useful for variants with lightning and weather changes. DTSBC - image features: Density Complexity + Size Complexity + Total Complexity + Balance Complexity; A - simple blockout; B - advanced blockout; C - models without materials; D - models with monochromatic materials; E - models with final materials; F - final level version; G - geometrical changes; L - lightening condition changes; W - weather changes; M - material changes; X - expression added; O - extra models added.

	Level Design Variant											
	A	B	C	D	E	F	G	L	W	M	X	O
DSTBC + Average Contrast	0,66	0,42	0,61	0,55	0,63	**0,73**	0,40	0,64	0,60	0,65	0,52	0,55
DSTBC + Average Contrast + Average Saturation + Hue Entropy	0,67	0,43	0,63	0,60	**0,72**	**0,81**	0,53	0,68	0,62	**0,75**	0,66	0,63
DSTBC + Average Contrast + Average Saturation + Hue Entropy + Luminosity Kurtosis	0,68	0,43	0,65	0,61	**0,82**	**0,82**	0,54	**0,70**	0,62	**0,75**	**0,71**	0,63
DSTBC + Hue Kurtosis + Hue Skewness + Saturation Entropy	0,63	0,47	0,62	0,59	**0,71**	**0,74**	0,46	0,55	**0,77**	0,68	0,52	0,57
DSTBC + Hue Kurtosis + Hue Skewness + Saturation Entropy + Luminosity Entropy	0,63	0,48	0,63	0,62	**0,71**	**0,77**	0,50	**0,77**	**0,80**	0,69	0,52	0,66

included in them (Fig. 3). These isolated opposite cases arise when one feature in a combination did not correlate individually. It can be observed in variant G (geometrical changes) where combined result of DSTBC + Average Contrast is equal to single Average Saturation correlation value (but with negative sign). On the other hand, the combined sets presented strong and very strong relationship for those level variants that for a single feature had only a few weak or moderate relationships: M (material changes), X (added expression) and O (extra models added). We can also observe that the more advanced level design stage (B to F) the stronger the correlation (Table 2). Even the worst level design variant for single feature - geometrical changes (G) - now shows significant strong positive relationship (Pearson $r = 0{,}54$ with $p < 0{,}01$ at best).

There is significant difference in correlation values for color-based features (color, luminance as well as descriptive statistics for HSL) between single feature correlation (Table 2) and combined value using those image features (Table 3). The single feature correlation values are rather small or even not significant on later design variants (G to O). However, when they are combined with other image features, they have shown the highest or the second-highest correlation value. This happens even if, for a given variant of the level design, a single color-

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_49

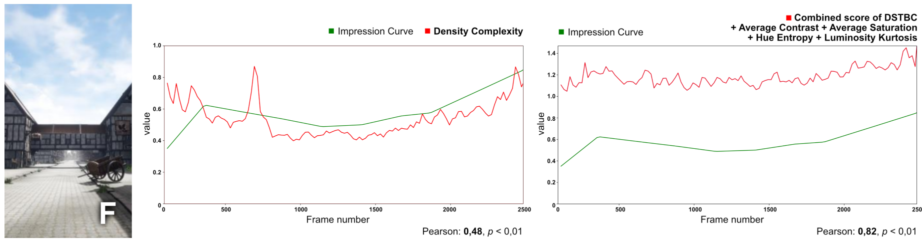

Fig. 3. Correlation plot examples for final level design variant (F variant, on the left). Two image feature correlation plots are presented: single image feature - Density Complexity (Pearson r = 0,48 with $p < 0{,}01$, center) and combined score of Density Complexity, Size Complexity, Total Complexity, Balance Complexity (DSTBC for short), Average Contrast, Average Saturation, Hue Entropy and Luminosity Kurtosis (Pearson r = 0,82 with $p < 0{,}01$, right). The combined score presents much higher correlation value than the component features separately with significant very strong positive relationship.

based feature did not show a correlation with the Impression Curve (mostly due to the high values of p).

4.3 Best features for different level design stages

Another aspect of the evaluation of the results was the changes of individual feature correlation at the subsequent stages of the game level design. Thanks to such approach, it was possible to assess the usefulness of the automatic evaluation method at different stages of the level design (from simple blockout with gray objects, trough materials and textures, to final design with lightning and atmospheric effects). The best image features or their combination to be used in such evaluation system will be the ones correlating regardless of the variant we are dealing with. The results for similar versions (such as first stage simple design been analyzed together) of the level were also compared. The image features with similar correlation values for each variant were considered the most promising. Such approach allowed us to eliminate those image features that correlated only in a single case.

For early stages of design that use blockout (A and B) we observed a significant correlation with the color-based image features, where at later stages (C to F) features from saliency and motion maps group showed better results (Table 2). What is more, most of the color-based image features tends to not show significant correlation at later stages, especially at the final level design. The exception here are the values of Hue Kurtosis, Hue Skewness and Luminosity Entropy for the weather change variant (W) with strong relationship. This showed that those image features could be added to the combined set to help verify how atmospheric effects affects users' impression of virtual space.

It is worth noticing that combination of HSL Entropy, Kurtosis and Skewness showed high or very high significant correlation results for variant the most

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_49

visually different from the rest - L - where lightning conditions are changed (day to night). For example, set combined of DSTBC + Hue Kurtosis + Hue Skewness + Saturation Entropy + Luminosity Entropy resulted for weather changes level design variant in very strong relationship (Pearson's r = 0,80, p < 0,01). Even this set is not universal for the whole process (other combined sets have given higher correlation results), it could improve Impression Curve estimation at design with rapid lightning or weather changes. At the same time, changes in lighting, weather or materials did not affect the shape of the Impression Curve, but did have a significant effect on the image features correlation.

5 Conclusion

The aim of this study was to investigate the existence of the correlation of the image features with the Impression Curve for game level design. The study shows that even a single image feature can describe the impression value with good precision (strong relationship, Pearson r > 0,5) for final level design. Best results were obtained by combining several image features using multiple regression (for image features: Density Complexity, Size Complexity, Total Complexity, Balance Complexity, Average Contrast, Average Saturation, Hue Entropy and Luminosity Kurtosis combined using multiple regression). Such set produced very strong positive relationship with Impression Curve values (Pearson r = 0,82 with p < 0,01 at best). What is more, significant correlation (strong to very strong) was observed regardless of level design variant, which makes it possible to apply image analysis at every stage of the level design process, making such solution more universal. The study also analyzed the possibility to use a different set of image features at different level design stages to get the highest results. The color-based image features were the best in this regard to be used at blockout stage of design (A and B, moderate to strong relationship) and HSL Entropy, Kurtosis and Skewness at stages with lightning and weather changes (L and W, moderate to strong relationship).

We saw many development opportunities for the idea of the automatic evaluation of game level design. The tests can be performed on production versions of game levels (taken from popular games), data about Impression Curve as well as image analysis could be obtained in real time or the study could be to extend with an Eye-tracker (to verify if there is a relation between the eye movement and Impression Curve). Also, the joined signals of EEG and Eye-tracker data can be analyzed as in [15]. Those four research ideas are carried out by us at the time of writing this article and the results will be published in the future. Improvements can be made in terms of calculation time as well, with a goal of real time analysis. For example, by applying faster classified like the one used in [12] for HUD detection, to obtain saliency maps in short time.

To sum up, the study has shown that Impression Curve value, and hence, the user impression of virtual 3D space, can be estimated with a high degree of certainty by automatic evaluation using image analysis of such level walkthrough. We propose usage of the combined image feature set for better estimation of

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_49

Impression Curve. For early stages of design (blockout and models without textures) different set can be used to increase the relationship strength.

References

1. Andrzejczak, J., Osowicz, M., Szrajber, R., Impression Curve as a New Tool in the Study of Visual Diversity of Computer Game Levels for Individual Phases of the Design Process. Computational Science - ICCS 2020, Springer, 2020
2. Dudek, P., Wang, B.: A Fast Self-tuning Background Subtraction Algorithm, IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 401-404, 2014
3. Jamil, N., Sembok, T. M. T., Bakar, Z. A.: Noise removal and enhancement of binary images using morphological operations, International Symposium on Information Technology, pp. 1-6, 2008
4. Milanfar, P.: A Tour of Modern Image Filtering: New Insights and Methods, Both Practical and Theoretical, IEEE Signal Processing Magazine, vol. 30, no. 1, pp. 106-128, 2013
5. Magel, K., Alemerien, K.: GUIEvaluator: A metric-tool for evaluating the complexity of graphical user interfaces. Proceedings of the International Conference on Software Engineering and Knowledge Engineering, SEKE, 2014
6. Peli, E.: Contrast in complex images, Journal of the Optical Society of America A, Optics, ImageScience, and Vision, 7, 2032–2040, 1990
7. Kumari, S., Vijay, R.: Image Quality Estimation by Entropy and Redundancy Calculation for Various Wavelet Families, International Journal of Computer Information Systems and Industrial Management Applications, Volume 4 pp. 027-034, 2012
8. Simons, D. J., Franconeri, S. L.: Moving and looming stimuli capture attention, Perception and Psychophysics volume 65, pp. 999–1010, 2003
9. Della-Bosca, D., Patterson, D., Roberts, S.: An analysis of game environments as measured by fractal complexity, In Proceedings of the Australasian Computer Science Week Multiconference (ACSW '17), Association for Computing Machinery, USA, Article 63, pp. 1–6., 2017
10. Kaehler, A., Bradski, G.: Learning OpenCV 3: Computer Vision in C++ with the OpenCV Library, O'Reilly Media, ISBN 9781491937969, 2016
11. Engineering Statistics Handbook, NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/, access: 02.02.2022
12. Kozłowski K., Korytkowski M. Szajerman D.: Visual analysis of computer game output video stream for gameplay metrics. Computational Science – ICCS 2020, LNCS, Vol. 12141, Springer, pp. 538-552, 2020
13. Lazar, J. et al., Research Methods in Human-Computer Interaction. Wiley Publishing, 2nd Edition, ISBN 9780128053904, 2017
14. Rogalska A., Napieralski P.: The visual attention saliency map for movie retrospection, Open Physics, vol. 16, no. 1, pp. 188-192, 2018
15. Szajerman, D., Napieralski, P., Lecointe, J.-P.: Joint analysis of simultaneous EEG and eye tracking data for video images, COMPEL - The international journal for computation and mathematics in electrical and electronic engineering, Vol. 37 No. 5, pp. 1870-1884, 2018
16. Ølsted, P. T., Ma, B., Risi, S.: Interactive Evolution of Levels for a Competitive Multiplayer FPS, Evolutionary Computation (CEC), 2015 IEEE Congress on, IEEE, pp. 1527-1534, 2015

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_49