VISUAL 2022

The Seventh International Conference on Applications and Systems of Visual Paradigms

ISBN: 978-1-61208-978-2

May 22nd –26th, 2022

Venice, Italy

VISUAL 2022 Editors

Gary Ushaw, Newcastle University, UK

VISUAL 2022

Forward

The Seventh International Conference on Applications and Systems of Visual Paradigms (VISUAL 2022) continued a series of events that put together complementary domains where visual approaches are considered in a synergetic view.

Visual paradigms were developed on the basis of understanding the brain's and eye's functions. They spread over computation, environment representation, autonomous devices, data presentation, and software/hardware approaches. The advent of Big Data, high speed images/camera, complexity and ubiquity of applications and services raises several requests on integrating visual-based solutions in cross-domain applications.

We take here the opportunity to warmly thank all the members of the VISUAL 2022 technical program committee, as well as all the reviewers. The creation of such a high-quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and effort to contribute to VISUAL 2022. We truly believe that, thanks to all these efforts, the final conference program consisted of top-quality contributions. We also thank the members of the VISUAL 2022 organizing committee for their help in handling the logistics of this event.

VISUAL 2022 Chairs

VISUAL 2022 Steering Committee

VISUAL 2022 Publicity Chairs

VISUAL 2022
Committee

VISUAL 2022 Steering Committee

Irene Yu-Hua Gu, Chalmers University of Technology, Sweden
Jinrong Xie, eBay Inc., USA
Marina Tropmann-Frick, Hamburg University of Applied Sciences, Germany
Sergio A. Velastin, Universidad Carlos III de Madrid, Spain

VISUAL 2022 Publicity Chairs

Mar Parra, Universitat Politècnica de València (UPV), Spain
Hannah Russell, Universitat Politècnica de València (UPV), Spain

VISUAL 2022 Technical Program Committee

Ana Paula Afonso, Universidade de Lisboa, Portugal
Marco Angelini, University of Rome "La Sapienza", Italy
Keivan Ansari, Institute for Color Science and Technology, Iran
Danilo Avola, Sapienza University, Italy
George Baciu, The Hong Kong Polytechnic University, Hong Kong
Hans-Peter Bischof, Rochester Institute of Technology, USA
Liudmila Bredikhina, Geneva University, Switzerland
Vittoria Bruni, Sapienza Rome University, Italy
Michael Burch, Eindhoven University of Technology, Netherlands
Patrick Callet, MINES ParisTech, France
Maria Beatriz Carmo, Universidade de Lisboa, Portugal
Satish Chand, Jawaharlal Nehru University, New Delhi, India
Haeyong Chung, University of Alabama in Huntsville, USA
Stefano Cirillo, Università degli Studi di Salerno, Italy
Gennaro Costagliola, University of Salerno, Italy
Tobias Czauderna, Monash University, Australia
Antonina Dattolo, University of Udine, Italy
José Mario De Martino, University of Campinas, Brazil
José Joaquim de Moura Ramos, University of A Coruña, Spain
Cosimo Distante, CNR / University of Salento, Italy
Pierre Drap, Aix-Marseille University, France
Omar Elharrouss, College of Engineering - Qatar University, Qatar
Ana Figueiras, Universidade Nova de Lisboa, Portugal
Valentina Franzoni, University of Perugia, Italy
Manuel González-Hidalgo, University of the Balearic Islands, Spain
Prashant Goswami, Blekinge Institute of Technology, Karlskrona, Sweden
Denis Gracanin, Virginia Tech, USA
Mustafa Hajij, Santa Clara University, USA
Wladyslaw Homenda, Warsaw University of Technology, Poland
Kamran Javed, Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Topi, Pakistan

Hamid Reza Karimi, Politecnico di Milano, Italy
Enkelejda Kasneci, University of Tübingen, Germany
Robert S. Laramee, University of Nottingham, UK
Nicola Lettieri, University of Salerno, Italy
Alexander Liebeskind, Columbia University, USA
Kari Lilja, Satakunta University of Applied Sciences | Tampere University of Technology | University Consortion Pori, Finland
Mei Kuan Lim, Monash University Malaysia, Malaysia
Magdalena Majdak, Institute of Polish Language - Polish Academy of Sciences, Poland
Alfredo Milani, University of Perugia, Italy
Bartolomeo Montrucchio, Politecnico di Torino, Italy
João Moura Pires, FCT NOVA, Portugal
Minoru Nakayama, Tokyo Institute of Technology, Japan
Umberto Nanni, University of Rome "La Sapienza", Italy
Paolo Nesi, University of Florence, Italy
Marian Sorin Nistor, Bundeswehr University Munich, Germany
Klimis Ntalianis, University of West Attica, Greece
Cristina Oliveira, Instituto de Ciências Matemáticas e de Computação | University of São Paulo, Brazil
Joanna Isabelle Olszewska, University of West Scotland, UK
Marcos Ortega Hortas, University of A Coruña, Spain
Francesco Pascale, University of Salerno, Italy
Janayna Passarinho, State University of Piauí, Brazil
Giuliana Ramella, National Research Council - CNR | Istitute for the Applications of Calculus "M. Picone" (IAC), Italy
Cleyton Luis Ramos Barbosa, Federal University of Pará, Brazil
Michele Risi, University of Salerno, Italy
Peter Rodgers, University of Kent, UK
Filip Sadlo, Heidelberg University, Germany
Ignacio Sanchez-Navarro, University of the West of Scotland, UK
Kristian Sandberg, Computational Solutions Inc., USA
Domenico Santaniello, University of Salerno, Italy
KC Santosh, University of South Dakota (USD), USA
Nickolas S. Sapidis, University of Western Macedonia, Greece
Antonio Sarasa Cabezuelo, Universidad Complutense de Madrid, Spain
Sonja Schimmler, Fraunhofer Institute for Open Communication Systems (Fraunhofer FOKUS), Berlin, Germany
Monica M. L. Sebillo, University of Salerno, Italy
Francesco Setti, University of Verona, Italy
Nada Sharaf, German University in Cairo, Egypt
Désiré Sidibe, Université Paris Sacaly, France
Gurjot Singh, Fairleigh Dickinson University, USA
Luciano Soares, Insper, Brazil
Nuno Miguel Soares Datia, Instituto Superior de Engenharia de Lisboa (ISEL), Portugal
Ryszard Tadeusiewicz, AGH University of Science and Technology, Poland
Diego Thomas, Kyushu University, Japan
Jean-Baptiste Thomas, NTNU, Norway
Marina Tropmann-Frick, University of Applied Sciences Hamburg, Germany
Gary Ushaw, Newcastle University, UK

Copyright Information

Table of Contents

A Collaborative Responsive and Fully Customizable System for Image Quality Assessment Based on Subjective Visual Perception

Maria Grazia Albanesi
Dept. of Electrical, Computer and Biomedical Engineering
University of Pavia
Pavia, Italy
Email: mariagrazia.albanesi@unipv.it

Abstract— **This paper describes a new system for image quality assessment related to subjective visual perception. The novelties of the contribution consist in the realization of a Web-based collaborative platform, which (a) allows a set of human observers to cooperate to the process of Mean Opinion Score generation in a complete asynchronous and distance-online participation, and (b) is fully responsive, thus the images can be tested on different devices, not only monitors or television sets, but also smartphones. A new method for the analysis of the results is also proposed; it is based on visual tables, which arrange the semantic of the experimental data in colored patterns, useful for understanding the properties of both images and observers. A set of target images has been identified and included in a Web based prototype to which testers can freely connect. The system is fully customizable because the code is freely available for download; in this way the programmer may easily change all the main parameters (images, number of impairments levels, type of impairment, and so on). Experimental results are reported and commented, with considerations on the possible applications of the system in different contexts.**

Keywords - subjective image quality assessment; visual analisys, collaborative Web-based process.

I. INTRODUCTION

Image quality assessment is an important task in modern multimedia processing, since the perceived quality can greatly influence the use of image information and the quality of experience [1-2]; moreover, the subjective evaluation of the perceived quality of a digital image is considered the most reliable method, since it can faithfully reflect the judgment of the human observer [3]. Unfortunately, organizing subjective image quality assessment experiments is very time consuming and requires a high economic effort and commitment. In this paper, a new system for subjective image quality assessment is proposed. The system was developed at the Department of Electrical, Computer and Biomedical Engineering (University of Pavia, Italy) and it has the main purpose of supporting scientists in the realization of experiments for subjective assessment of the quality of digital images. The most used approach (in the following, the *traditional approach*) in literature is to carry out the evaluation of the perceived quality experimentally using a set of human observers, which are asked to rate the perceived quality in different situations. In this process, there are several issues that matter:

- The choice of human observers is particularly critical, since they must be sufficiently educated on the standardized protocol to be used in the experimental evaluation and must also have peculiar characteristics, e.g., normal or corrected to normal visual acuity, a good sensitivity to colors, a good ability to maintain high attention and concentration as the experiments can also last several minutes.
- The realization of the experiments for perceptive evaluation of image quality in the laboratory are often too expensive and time consuming. The setup of the necessary hardware can involve the use of a high volume of resources and efforts.
- The choice of the set of images to be evaluated can also greatly influence the results of the experiments. Generally, the choice is made according to the type of impairment to consider. There are many datasets of digital images freely downloadable to be used in the experiments, for example the very well-known databases by the Laboratory for Image & Video Engineering (LIVE) [4].

For many years, the subjective assessments of image quality have been realized following the traditional approach, which has been particularly relevant in some applications, e.g., in television broadcasting [5]. Since 2009, some innovative proposals began to appear [6-8], with the intent to replace or to be alternative to the traditional method. They were based on a *crowdsourcing approach*, namely the possibility of recruiting observers on the Web and the realizations of the experiments using a Web-interface. This approach has several advantages and disadvantages, which must be properly balanced. The crowdsourcing approach reduces costs, as it is not necessary to set up an effective evaluation environment in the laboratory (hardware set up and staff commitment). Furthermore, the recruitment phase of the observers can also be delegated to another subject (e. g., an external company, as in [8]) or using collaborative platforms, i.e., social networks. The main disadvantage of this approach is that you lose the control over the choice of observers and the standardization of the method. For this reason, it is very important to provide observers with a precise and detailed protocol with all the instructions they must follow, including those relating to the implementation of the experiment, for example the distance from the monitor, lighting, or other details. However, there is no guarantee that the observer follows the protocol correctly as the process of expressing the opinion on the perceived quality is not supervised. Although the last aspect seems particularly critical, the results in the literature have been encouraging in identifying the crowdsourcing approach as very similar if not equivalent to the traditional approach, by comparing the Mean Opinion Score (MOS) in both the approaches on the same target dataset, with a high correlation coefficients and confidence interval [8].

The main goals of the project here reported try to overcome some of the problems of both the traditional and the crowdsourcing approaches, by proposing a hybrid system. Moreover, an innovative tool for analyzing human observer reliability is described, to fill the gap in literature on the critical issue of the selection of subjects involved in the experiments.

The rest of the paper is organized as follows: Section II describes the goals addressed and the method implemented in the system. Section III describes the experiment design and analysis. Section IV provides some further ideas for using the system in different applications. The acknowledgement and conclusions close the article.

II. Goals and Methods of System Implementation

The proposed system is a hybrid between the traditional approach and the crowdsourcing approach. It was called Subjective Image Quality Assessment (SIQA) project [9] and it is based on a Web interface for experimental evaluation of the perceived quality. However, the possibility of downloading the entire project and installing a stand-alone configuration (still maintaining the web interface) allows to use it also for the traditional approach. We called it a hybrid system because it clearly can be exploited while retaining the positive aspects of both solutions, e.g., the traditional and the crowdsourcing approach. In fact, there are two methods for using the SIQA system:

- Offline method: the open-source code [10], was developed in Phyton, under Massachusetts Institute of Technology (MIT) License; it can be freely downloaded and used as it is or modified to adapt the experiments to your commitments. This is like the traditional approach, but for the user interface (a web browser) and the image dataset (freely downloadable o replicable). This method can also guarantee the important aspect of customization (see Section II.A, Innovations of the System).
- Live method: by taking the “Live Test” on the SIQA website [9]. This is performed on a subset of images and, clearly, it follows a pure crowdsourcing approach. However, it can be used for the important phases of the traditional approach of *training* the observers and their *warming-up* at the beginning of the subjective quality assessment experiments [3]. Therefore, the system is useful regardless of which approach (traditional or crowdsourcing) you use and can be adapted to different experimental contexts.

The two operating methods also guarantee some innovative aspects, with respects to the current state-of-the-art, which are detailed in the next paragraph.

A. *Innovations of the System*

Using the SIQA system in one or both the methods before described, you can have some interesting and innovative advantages:

- You can decide to have full control over the choice of observers, as in the traditional approach, or to relax it (crowdsourcing approach).
- You can choose one of the approach or both according to your needs: for example, you could choose the crowdsourcing approach for a test phase or for the selection of observers and then choose a more traditional approach to improve the standardization of the process and obtain the real experimental data.
- The SIQA Web site is fully responsive, so it can be used to test human observer perception on different kind of devices, including smartphones. Evaluating user experience on smartphone has become more and more important over the years [11].
- The possibility of completely downloading all the code and being able to modify it allows the scientist to easily replace the image datasets (taking it form the many available in literature [4] or a personalized dataset) and modify their characteristics, for example the number of levels and type of impairment, the type of images (gray-level or color images, resolution, format, or compression standards).
- The possibility of changing the quality impairment is very important: it can be not only noise, compression, or other engineering parameters, but also coming from other disciplines, to extend the concept of *image quality* to other fields, for example psychology, which is interesting in communication and advertisement (see Section IV, Examples of Application of the System).

Besides the SIQA system, in the paper an innovative tool, called *Visual MOS Table*, for analyzing the experimental data is proposed and described. It is a visual tool: by arranging the MOS values in a bidimensional table and with a simple but effective use of colors, it is very simple to analyze which images in the dataset can be considered “outliers” in the perceptual process. Moreover, the *Visual MOS Table* can be used also to identify which human observers are too strict, or too permissive, in assessing quality.

B. *Impression or quality?*

At first glance, SIQA project may seem very similar to systems used for user impression evaluation, such as the so popular solutions of photo sharing on Web. However, there are some substantial differences: first, in a subjective quality assessment task (as in SIQA project), opinion score is statistically evaluated not only for what concern the average, but also the standard deviation, which gives relevant information about the homogeneity of the evaluation task.

Moreover, in a real subjective evaluation experiment, the user does not choose the image to rank, and the set of assessed images is the same for all the subjects. These constraints do not apply to standard photo impression evaluation. For this reason, SIQA project evaluate the subjective quality (as defined in literature), rather than a simple user impression on the Web.

III. EXPERIMENT DESIGN AND ANALISYS

The system is fully compatible with the method recognized in the literature among the best ones for subjective evaluation, i.e., the single stimulus Absolutely Category Rating (ACR) test [5]; in fact, in literature, there is no evidence that a double stimulus method is more accurate than a single stimulus [12]. Moreover, the single ACR stimulus is one of the best in optimizing time effort of the experiments, especially if we need to put a temporal limit to each experimental session (in general no more than 30 minutes). The single stimulus ACR method ask the observers their opinion about the perceived quality of only one image at time, by expressing a rank according to five scores, from 1 to 5, according to a well-known equivalence, shown in Table I. The values of the scores are collected on a pool of images and for a set on human observer, and the standard results of MOS and its Standard Deviation (SDOS) are provided.

The SIQA system has been tested first in a set of closed experiments (i.e., inside the laboratory) on different kinds of images (color, grey levels) and different visual impairments (lossy compression, noise addition, blurring) of increasing severity. On the SIQA Web site [9] the results of these preliminary tests are available by clicking on the second button (i.e., "Download presentations") in the section Download (see Figure 1).

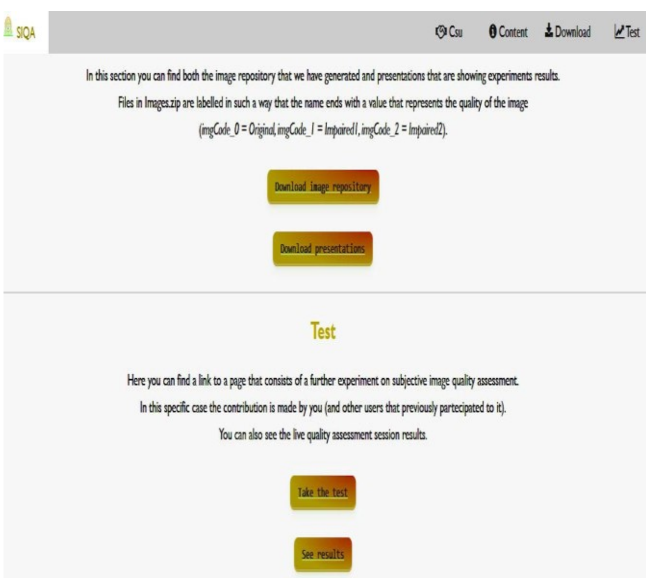

Figure 1. The Donwload Section (on the top) and the Live Test Section (on the bottom) of the SIQA Website (https://siqa.pythonanywhere.com).

TABLE I. THE FIVE SCORES OF THE ACR METHOD

Score	Perceived Quality	Perceived Impairment	Colors in Visual MOS Tables[a]
5	Excellent	Imperceptible	Green
4	Good	Just Perceptible but not annoying	Light Green
3	Fair	Perceptible and slightly annoying	Yellow
2	Poor	Annoying	Light Red
1	Bad	Very annoying	Red

a. The meaning of this field is explained in Section III.A

As explained in the previous paragraph on innovations, there are two design methods: Live (online) Test and Offline test. The first correspond to the crowdsourcing approach, the second to the traditional one. In the Live Test is available by connecting to the SIQA Website and by clicking on the button "Take the test". The test is performed using a pre-defined set of three different versions of ten color images: "original", "impairment1" and "impairment2". The ten original images are depicted in Figure 2. The impaired versions are obtained from the ten original ones, by applying the standard Joint Photographic Experts Group (JPEG) compression algorithm, with different quality levels. Degradation on impairment1 images should be less visually noticeable than the one on the impairment2 images.

Figure 3 shows a moment of the score expression in the interface of SIQA. The score is requested by a simple mouse click; other interfaces, for example in [13], store the score using a visual stimulus recording, by an eye tracker. However, we have decided the simplest mouse-driven interface because we want to assure the possibility to use the Live Test also on other devices, e.g., smartphone (where mouse click is substituted by a simple touch).

The entire test on the thirty images takes approximately ten minutes. We have chosen this duration because it can be easily used to select the observers in trial test or for the reheat phase in traditional experiments. The strictly controlled conditions of traditional test are replaced with simple instructions to the observer:

- Observe all the images under similar conditions: (viewing distance, brightness, artificial lighting, screen contrast).
- Give an evaluation for each image in a maximum time of 15 seconds.
- Wear contact lenses or prescription glasses, if any.
- Take the test only once.

After the test, the scores expressed by the online observer are added to the set of previously stored scores in the system, and MOS and SDOS are updated. They plots are available by clicking on the button "See Results". For brevity, only the MOS experimental data are reported and analyzed in the paper (SDOS values are available online).

At the time of writing this paper, the values of MOS for each image and each level of impairment are the ones shown in Figure 4: on the x-axis there are the ten images of Figure 2. The three MOS plots refer to the different levels of JPEG quality (Original, Impairment1 and Impairment2). The trend

 ISBN: 978-1-61208-978-2

of MOS is perfectly consistent with the levels of impairment ("original MOS" is always over the impairment1, which is always over impairment2). Only for one image (i.e., image "tree", showing section of a tree trunk, see Figure 2g) the three values are very closed: 3.79, 3.66, and 3.37, for, respectively, the original, impairment1, and impairment2 image. A possible explanation of this fact is that probably the radial symmetry of the texture of image "tree" does not allow users to clearly perceive the block distortion, which is typical for JPEG compression.

Figure 2. The ten images for the Live Test: (a) Water (b) Books (c) Hand (d) Dogs (e) Tiger (f) Mountains (g) Tree (h) London (j) Poppies (k) Bear.

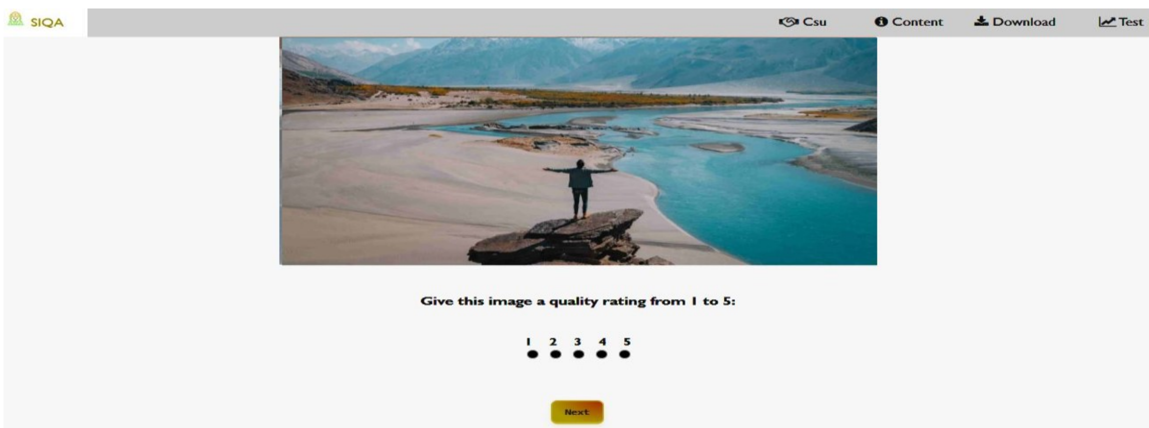

Figure 3. Example of the opinion expression for image "Mountains" in the Live Test.

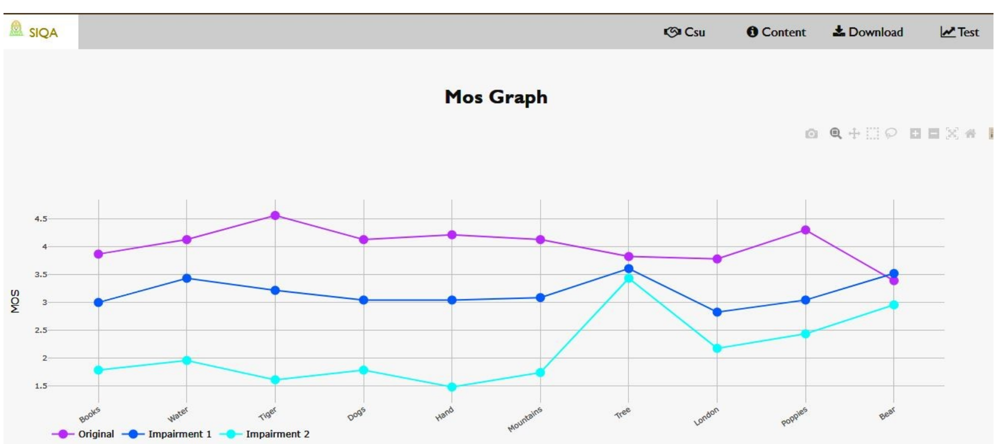

Figure 4. The MOS computed in the Live Test for the ten images of Figure 2.

The second design method is the Offline test. On the SIQA Web site [9] a set of 135 preloaded images (45 original and 90 impaired versions) are available by clicking on the first button (i.e., "Download image repository") in the section Download. In the following paragraph, a discussion of the analysis of the MOS experimental data are discussed to show the application of the visual tool called *Visual MOS Table*.

A. A new tool for experiment analysis: the Visual MOS Table

The image repository for Offline method refers to gray-level images; Figure 5 shows the first ten (original) images. We have used them in an experiment of subjective image quality assessment to show the utility of *Visual MOS tables*.

The experiments consider three levels of quality, i.e., Unimpaired (e.g., Original), Slightly impaired (JPEG lossy compression, quality factor 25) and Highly impaired (JPEG lossy compression, quality factor 12). The MOS has been measured for 13 expert observers (Evaluator 1-13) and 8 non-expert observers (Evaluator 14-21). Figure 6 shows the *Visual MOS Table*, created on the MOS data collected in the experiments. The table cells are colored to reflect the MOS values: green and light green for MOS 5 and 4, yellow for MOS 3, light red and red for MOS 2 and 1. Data are arranged in two dimensions: the first refers to the columns, and columns containing data of the same level of quality define an *Area of quality level*. In Figure 6, the *Area of quality level* "Unimpaired" is highlighted with a blue box. By analyzing an *Area of quality level* across the columns, it is very fast and easy to identify images which are outliers in the process of subjective quality assessment. For example, in the *Area of quality level* "Unimpaired" yellow and red (light and dark red) cells refer to outlier images, i.e., images that have achieved a different score than one would expect. In Figure 6, for expert observers, image "Wuhan", "Baby" and "Building" have collected five and four outliers (over 13 observers). This suggest that they are the most difficult images to rank correctly.

The *Visual MOS Table* can be analyzed also in the other dimension, by row. In this case the Area refer to homogeneous experimental constraints *(Area of constraint)*. In the example, we distinguish between expert vs. non-expert observers and in Figure 6 the *Area of constraint* "non-expert" is identified by a purple horizontal box. In this case, it is possible to identify human observers who are not well suited to effectively carry out perceptual judgment. Therefore, the system is useful not only to assess image quality, but also "tester (human observer) quality". For example, among the non-expert area, the Evaluator n. 16 has judged almost every image with very high opinion score (all the cells are green, independently on the real quality!). By analyzing the *Visual MOS Table*, we also discover that among experts, the Evaluator n. 12 was unable to distinguish between images at the highest quality level (Unimpaired) and that of intermediate quality (Slightly Impaired), since all the cells in the line "Evaluator n. 12" are green in the two *Areas of quality*. Therefore, Evaluator 12 and 16 could be considered outliers and could be discarded by the pool of observer in a next experiment. We suggest using the *Visual MOS Table* also to assess the "quality" of the human observers. The choice of observers is a very crucial point for the success of the subjective evaluation of perceived quality experiments and the *Visual MOS Table* can be a valid help to assist this task.

IV. EXAMPLES OF APPLICATIONS OF THE SYSTEM

The system can be customized by downloading the code and modifying the parameters, for example the number of impairment levels or the constraint conditions. These customizations reflect on the *Visual MOS Table* by adding further *Areas of quality* or *Areas of constraints*. For example, instead of using the two constraints expert vs. non-expert observers, constraints may be changed in "observing on the monitor" and "observing on the smartphones", if the goal of the experiment is to investigate the impact on perceived quality on different screen.

Another interesting application is to the definition of a more extensive concept of *quality*; if only one *Area of quality* is used, observers could be asked to evaluate not so much the perceived quality at the increasing levels, but the effectiveness in the transmission of a semantic message. Suppose for example to have five images and that the task is to evaluate how they are perceived by the humans for what concerns the efficacy in an advertisement message. By analyzing the Visual MOS Tables you can easily find which is the best image which obtained the highest score (i.e., the highest number of "green" or light green cells), possibly studying different viewing hardware (television or smartphones).

V. CONCLUSION AND FUTURE WORK

The paper describes a hybrid system for both traditional and crowdsourcing subjective assessment of image quality using a classical single stimulus ACR method. An innovative visual tool has been also described for the analysis of MOS experimental data, both for finding outliers in the image set or in the human observers. Future work will try to adapt the SIQA system to: (a) evaluation of video quality and (b) to rehabilitation of patients with cognitive deficits. In the latter case, instead of measuring the perceived quality, the human observer would be asked to express a judgment on the semantic of image, to check the ability of the patient to recognize different targets (faces, situations, environments).

ACKNOWLEDGMENT

The author would like to thank very much the following computer engineers of the staff of the Digital Content Retrieval Lab (University of Pavia) for the collection of the images and the deployment on the Web of the system and of the downloadable code: A. Al Masoud, F. Amato, A. S. Blindu, M. Bottini, D. P. Lotito, F. Marinelli, D. Murer, R. Petri, and F. Tagliani.

REFERENCES

[1] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," IEEE Trans. Image Process., vol. 15, no. 2, pp. 430–444, Jan. 2006.

[2] Z. Bovik, Modern Image Quality Assessment, Morgan & Claypool Publisher, 2006.

[3] H. R. Hu and K. R. Rao, Digital Video Image Quality and Perceptual, Taylor and Francis, 2006.

[4] Laboratory for Image and Video Engineering, University of Texas, USA (LIVE) *Subjective Image Quality Database.* [retrieved: April 2022]: https://live.ece.utexas.edu/research/quality/subjective.htm

[5] "Methodology for the subjective assessment of the quality of television pictures," ITU-R Recommendation BT.500-12, Sept. 2009.

[6] K. T. Chen, C. C. Wu, Y. C. Chang, and C. L. Lei, "A crowdsourceable QoE evaluation framework for multimedia content," in Proc. ACM Multimedia 2009. ACM, pp. 491–500, 2009.

[7] Z. Wang, D. Tao, and P. Liu, "Development and Challenges of Crowdsourcing Quality of Experience Evaluation for Multimedia", Wang Y., Xiong H., Argamon S., Li X., Li J. (eds) Big Data Computing and Communications. Lecture Notes in Computer Science, vol 9196. Springer, 2015: doi: https://doi.org/10.1007/978-3-319-22047-5_36

[8] F. Ribeiro, D. Florencio, and V. Nascimento, "Crowdsourcing subjective image quality evaluation," Proc. of 18th IEEE International Conference on Image Processing, IEEE Press, Sept. 2011, pp. 3158-3161.

[9] University of Pavia. *Subjective Image Quality Assessment.* [retrieved: April 2022]: https://siqa.pythonanywhere.com/

[10] University of Pavia. *SIQA Software*. [retrieved: April 2022]: https://github.com/aiman-al-masoud/image_quality_assessment

[11] Perceived Image Quality on Mobile Phones with Different Screen Resolution, doi: https://doi.org/10.1155/2016/9621925

[12] R. K. Mantiuk1, A. Tomaszewska. and R. Mantiuk, "Comparison of four subjective methods for image quality assessment", COMPUTER GRAPHICS Forum, Vol. 31, number 8, pp. 2478–2491, 2012: doi: https://doi.org/10.1111/j.1467-8659.2012.03188.x

[13] O. Le Meur, A. Ninassi, P. Le Callet, and D. Barba, "Overt Visual Attention for Free-viewing and Quality Assessment Tasks: Impact of the Regions of Interest on a Video Quality Metric", Signal Processing: Image Communication, Vol. 25, Issue 7, pp. 547-558, Aug. 2010, ISSN 0923-5965, doi: https://doi.org/10.1016/j.image.2010.05.006.

Figure 5. The first ten images of the downloadable dataset; in figure, only the original unimpaired version are shown.

	Unimpaired images										Slightly impaired images										Highly impaired images									
	01 - Wheel	02 - Boat	03 - Rides	04 - Wuhan	05 - Guy	06 - Beach	07 - Car	08 - Pedestrians	09 - Baby	10 - Building	01 - Wheel	02 - Boat	03 - Rides	04 - Wuhan	05 - Guy	06 - Beach	07 - Car	08 - Pedestrians	09 - Baby	10 - Building	01 - Wheel	02 - Boat	03 - Rides	04 - Wuhan	05 - Guy	06 - Beach	07 - Car	08 - Pedestrians	09 - Baby	10 - Building
	"**Expert**" evaluators group																													
Evaluator No. 1	5	4	3	5	5	4	4	5	3	5	3	3	2	5	3	4	3	3	3	4	2	2	1	3	2	1	1	2	1	2
Evaluator No. 2	4	5	5	5	4	5	2	5	3	5	3	5	3	5	4	5	2	4	5	5	1	2	1	4	5	2	1	5	3	5
Evaluator No. 3	5	5	5	3	5	5	5	5	4	5	3	4	3	2	3	2	3	4	3	4	2	1	2	2	2	2	1	3	4	4
Evaluator No. 4	5	5	4	4	4	5	5	5	4	3	3	5	2	2	2	4	4	3	2	3	1	1	1	3	1	2	1	1	2	2
Evaluator No. 5	5	5	5	5	5	5	5	5	3	5	2	5	3	3	5	3	4	5	5	3	1	1	1	3	3	1	1	3	4	1
Evaluator No. 6	3	4	5	4	4	4	5	5	5	4	3	3	2	3	3	3	3	3	3	3	2	2	2	2	2	2	1	2	2	2
Evaluator No. 7	5	5	5	3	4	4	4	5	4	3	2	3	2	2	2	1	2	3	3	3	1	1	1	2	1	1	1	1	1	1
Evaluator No. 8	2	4	5	5	4	3	5	4	5	4	2	2	2	3	2	2	3	3	4	2	1	1	1	1	1	1	1	2	1	1
Evaluator No. 9	5	4	5	1	5	5	1	5	5	4	3	4	2	3	4	2	3	2	3	3	3	2	1	5	1	2	1	1	2	1
Evaluator No. 10	5	4	4	3	4	4	5	4	4	3	2	3	2	4	3	3	3	3	4	3	1	1	1	2	2	1	1	2	2	2
Evaluator No. 11	5	5	5	4	5	5	4	5	5	3	3	3	3	3	2	3	3	3	3	4	1	1	1	3	1	2	1	2	2	3
Evaluator No. 12	5	5	5	4	5	4	5	5	5	5	4	5	4	5	4	4	5	5	5	4	2	2	2	5	5	2	3	4	4	5
Evaluator No. 13	5	4	3	3	3	4	3	4	3	4	2	2	2	2	2	2	2	3	2	2	1	2	2	2	1	1	1	1	1	2
SD of MOS for "**Expert**" evaluators group	1,0	0,5	0,8	1,2	0,7	0,7	1,3	0,4	0,9	0,9	0,6	1,1	0,7	1,2	1,0	1,1	0,9	0,9	1,1	0,9	0,7	0,5	0,5	1,2	1,4	0,5	0,6	1,2	1,2	1,4
	"**Non-expert**" evaluators group																													
Evaluator No. 14	3	4	5	5	2	3	5	4	3	5	4	4	4	4	3	3	4	2	3	3	3	4	2	3	1	2	2	3	2	4
Evaluator No. 15	4	4	4	5	4	4	5	5	5	4	5	2	2	3	3	3	3	4	4	5	1	1	1	3	1	1	1	3	3	3
Evaluator No. 16	5	5	5	5	5	5	5	4	5	5	5	5	5	5	5	5	5	4	5	5	5	5	5	5	5	5	5	4	5	5
Evaluator No. 17	4	4	4	5	4	3	5	5	4	4	4	4	4	3	4	4	2	4	3	5	3	2	1	2	1	1	2	1	2	2
Evaluator No. 18	4	4	4	3	4	3	3	3	4	4	3	3	2	1	2	2	1	3	3	4	1	1	2	2	1	1	1	2	2	1
Evaluator No. 19	5	5	3	4	5	3	4	3	3	2	3	4	2	3	4	3	3	4	2	4	2	2	1	3	2	2	1	2	2	2
Evaluator No. 20	5	5	5	5	5	5	5	5	5	5	4	3	3	5	4	3	3	3	3	3	1	1	1	3	1	1	2	3	3	1
Evaluator No. 21	3	3	3	4	4	3	3	4	4	3	2	2	2	3	2	2	3	3	3	2	1	1	1	2	1	1	2	1	2	2
SD of MOS for "**Non-expert**" evaluators group	0,8	0,7	0,8	0,8	1,0	0,9	0,9	0,8	0,8	1,1	1,0	1,1	1,2	1,3	1,1	1,0	1,2	0,7	0,9	1,1	1,5	1,6	1,4	1,0	1,4	1,4	1,3	1,1	1,1	1,4
Global SD of MOS	0,9	0,6	0,8	1,1	0,8	0,8	1,2	0,7	0,8	0,9	0,9	1,1	0,9	1,2	1,0	1,0	1,0	0,8	1,0	1,0	1,1	1,1	0,9	1,1	1,4	0,9	1,0	1,1	1,1	1,4
	0,9										1,0										1,2									

Figure 6. The *Visual MOS Table* for the experiments of subjective quality assessment on the first ten images of the downloadable set (see Figure 5). The blue and purple rectangles define the *Area of quality level* and the *Area of constraint* of the table, respectively.

Review of Core Graphic Design Principles Used in Computer Games

An Evaluation of Graphical User Interface Design

Christy Martin, Michael Cooper
Visual Communication Studio + Consultancy
Queensland,
Australia
e-mail: voxelandpixel@gmail.com

Manolya Kavakli
AIE Institute
33 Mountain St, Ultimo
Sydney, Australia
e-mail: manolya.kavakli@aieinstitute.edu.au

***Abstract*—Our goal in this paper is to review the use of core graphic design principles in computer games. We believe that the architecture of great Graphical User Interface (GUI) design is built upon a solid foundation of core graphic design principles. First, we define core principles used in graphic design, such as the grid, hierarchy, scale, balance, and framing. Then, we analyse the examples of core graphic design principles in computer games. Our findings state that there is evidence of all being utilised, despite the disparity in game genre.**

Keywords- computer games; GUI design; graphic design principles; user interface; evaluation.

I. INTRODUCTION

Our current technologies for visual communication (through print, digital, environment and increasingly virtual/interactive means) all fall back on an established set of basic design principles to consider when approaching a design, across these platforms [1][2]. New technologies are expanding considerations needed for graphic design layout, with screen sizing and touch capabilities also impacting possibilities in design. There are opinion articles about best user interfaces as seen in games [3][4] and what the medium is capable of delivering on, however no reviews about how basic graphic design principles are used in interface design.

The core principles of graphic design have remained approximately the same from modernist-era developments, having proven their efficacy over time. We have a basic framework for graphic design, developed and expanded upon since the International Typographic Style (ITS) emerged in the 1960's, to assist us in organising and understanding the digital era challenges we now face as designers, in a fast-moving and ever-more technical world. Our potential toolkit for executing graphic design grows, as we experiment with new technologies, and engage with new ideas and social/political constructs.

Design principles prove to offer a great foundational architecture for visual communicators to build their designs, with huge scope for further discussion and development of graphic design concepts within the bounds of the two, three or even fourth dimensional spaces we interact with today.

Graphic design serves as the bridge between the game software and the player. Graphic designers design the different User Interfaces (UI) the player interacts with. The menus, icons, navigation bars, data, and in-game displays must be organized with the user utmost in mind, and simple to understand. They must accommodate users from different backgrounds and cultures by applying universal design principles to communicate in the most efficient way possible.

The rest of this paper is organized as follows. Section II describes the core graphic design principles. Section III demonstrates the use of core graphic design principles in computer games. Section IV addresses the analysis of underutilized elements of graphic design in contemporary computer games. Section V draws an evaluation with respect to the GUI design and lessons learnt from the analysis. The acknowledgement and conclusions close the article.

II. CORE GRAPHIC DESIGN PRINCIPLES

The architecture of great graphic design is built upon a strong foundation of core graphic design principles that most designers begin with [2]:

- the grid
- hierarchy
- scale
- balance (symmetry and asymmetry)
- framing

These five principles give structure to layout, to allow the designer to then experiment with other elements of form, for instance: colour, line, pattern, texture, and layers.

A. The Grid

A grid could be seen as the "bones" of a design layout - a series of columns with margins, to place text and image within – the organization of content. A grid allows the designer to align content vertically and horizontally across multiple pages of layout. Column width and depth can either be set as a strict adherence or varied to the designer's specifications, making this structural element work hard throughout the design (thus giving flow and balance and allowing great control over hierarchy).

From this foundation, the designer may choose to break out of the grid boundaries, to enhance the impact of the work; with Gestalt principles, asymmetry, contrast, and scale contributing to this dynamism.

 ISBN: 978-1-61208-978-2

B. Hierarchy

When we speak of hierarchy as a graphic design term, we are insinuating an order of importance in the reading of information, the visual hierarchy of design. Creating a hierarchy of visuals aims to clearly convey a message, a goal achieved using a range of concepts defined as the Basic Design Principles. Sources might list slightly different versions of the "Principles of Graphic Design", such as in [5]. The terminology used might be different, but the fundamental concepts are the same.

The gaze concept draws on psychological studies to determine the way we take in valuable information – how we process what we see, what we look for in (for instance) a piece of design work. Our viewing patterns tend toward creating order to conceptualise, rationalise and group what we see; a factor further explored in Gestalt principles. We are looking for things which stand out to guide us visually: differentiation.

The message or information - the communication in design - is often the order of text content, creating typographic hierarchy to clearly, as Ellen Lupton [5] states, "add difference methodically" and create visual "signals of difference, signals of separation".

In layout, defining and differentiating content, such as heading, sub-heading, body text and caption can be achieved by adjusting the character of the text, using combinations of:

- size and scale (drama/dynamism)
- weight (bold, medium, light, italic)
- colour (emotion, tone)
- spacing (flow and pause between text)
- case (all capitals, all lowercase)
- character (typeface selection).

Hierarchy provides order and direction to structure. Our eye naturally gravitates to differentiations, spatially and typographically, so with appropriate design we can communicate an order of importance in our messaging and lead a user's eye around a page or screen.

Typographic hierarchy is particularly important to communication as the ability to clearly read and understand where information begins, and ends, gives the reader comfortable flow (with visual pauses and punctations) when absorbing the text.

Hierarchy also makes life easier for the designer, as we can set our particular "styles" for typography, line spacing and kerning at the beginning of the design process, and in layout software programs like InDesign, apply them across the board.

C. Scale

Scale gives dimension to design. Designers use contrast in scale to draw the eye, for impact and added drama. Scale of text or image on the page can assist with hierarchy. Appropriate scale for text sets a comfortable reading size, or makes a headline stand out. In design, we see two types of scale:

1. **Objective scale:** how large something is (in reality, or the scaled down actuality of the real thing).
2. **Subjective scale:** how big we think something is (our impression; in relation to other things).

Similar sized graphics in a design can lead to flatness, or monotony, therefore adding variation through scale creates movement and flow.

D. Balance (symmetry and asymmetry)

A graphic designer will often play with the shapes and text in their design piece until a tenuous balance, or flow, develops. The resulting rhythm and balance, contribute to the overall feel of a piece of design. The goal of both symmetrical and asymmetrical design principles is to engender a sense of harmony and balance to a piece - a stability of form and flow.

Whilst symmetry harnesses the power of harmony and simplicity, asymmetry challenges the eye with form which may appear random, however placement of elements is deliberate, to add interest and movement. Symmetry through simplicity of form gives us strong, clear messaging. Asymmetry through a harmonious balance guides us through the communication, promoting interaction.

E. Framing

Framing concentrates focus within the design space. The edges of the page, screen or object form a natural barrier. Framing for images through cropping is a subtle skill; the challenge being to identify and choose an appropriate section, angle, and subject. Margins on layouts also provide a framing opportunity. The placement or orientation of design elements on a page become dictated by the grid, with text, image and illustration layered across these lines, but all have an end-point – the edge of the page, or an internal shape.

Framing "creates the conditions for understanding an image or object", according to Lupton & Phillips [2, pp.116). Content within a frame becomes our focus, the frame or border itself a tool of visual acuity.

All graphic design is contained in a frame of reference, be it a device screen, a book jacket, magazine spread or package. The way a graphic designer chooses to use the frame is where creativity surfaces - sometimes what is left out of the frame can be as important as what remains inside.

III. USE OF GRAPHIC DESIGN PRINCIPLES IN GAMES

The GUI is the arrangement of visual components that act as a means of communication between the user and various aspects of the game code, for example, in-game interactivity, information displays, narrative, and settings. A graphic designer designs the user interface (UI) of a game. A single game often has multiple user interfaces; thus, the designer decides the placement of all these visual elements and typography to create not only a pleasing aesthetic, but also to avoid restricting gameplay. There are a few studies published using design principles to enhance engagement in therapy-based video games [22] and cloud games [23].

UI in games takes on different forms (Figure 1), such as **diegetic** (elements that exist within the game world), or **non-diegetic** (an on-screen UI element that is separate from the game world) [7]. These two approaches can markedly affect the way the graphic design appears in the game, but the core design principles in creating it will remain similar. Other UI types are **Meta** (elements which exist within the game world and are used as a form of UI) and **Spatial** (an element or pop-up which appears in the game but is not of the game world).

Figure 1. Table of GUI types and uses [7].

In this paper, we will examine the 5 Core Graphic Design Principles of Grid, Hierarchy, Scale, Balance and Framing as they apply to recent game titles with a range of GUI styles.

A. *The Grid*

There is an acceptable level of valuable screen 'real estate' used for UI, which allows it to be visible and playable, but not overtly intrusive. Perspective and view (keeping it clear) are of utmost importance.

Use of a grid to develop the staging for a game's UI can depend on the style of game (the difference between a console game and a mobile game, for example, may include the screen size and shape, style of game, and simplicity of controls).

GUI for diegetic games operates on a slightly different basis to those which present UI as a screen overlay. Diegetic UI will often be gridded out, with the same hierarchical principles, but when they appear onscreen diverge; the viewpoint of the player affects how the UI appears (dependent on the element in question and camera view being used, the grid is skewed.).

The non-diegetic grid remains flat and neat, arranged around the screen (a 'Head's Up Display' or 'HUD'), or popping up as a layered sub-screen when required, with ease of use and clarity the goal.

Appropriate design of UI grid systems will give the user a sense of space - that all elements in the layout have the right amount of proximity. In this example from first-person shooter (FPS) "Destiny 2" (Bungie Games), we see use of grid systems to position inventory UI onscreen, in squares with strict alignments, hanging from and ending at a specified point (Figure 2).

Diegetic UI, such as in the action-adventure, third-person shooter "Tom Clancy's The Division 2" (Ubisoft), require the same build structures but appears onscreen quite differently, often from the perspective of the character (Figure 3). The information and icons are in columns, with necessary alignment and structure. The player views them as 'built in' to the world, in this case as a projection from the character's watch.

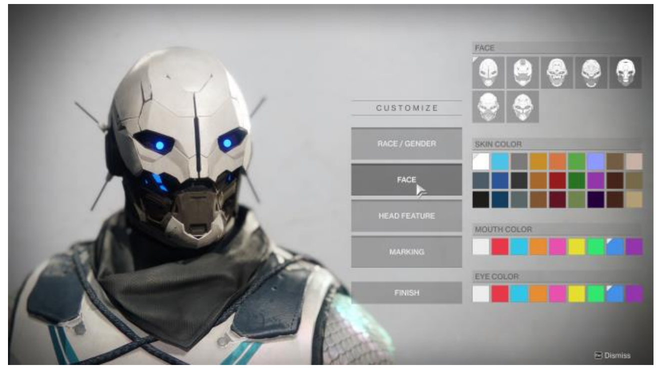

Figure 2. Example of grid elements in non-diegetic UI ("Destiny2", Bungie Games) [8].

"Tom Clancy's The Division 2" gives us a good example of a UI where the character is looking at an 'in-world' object and interacting with it. We can clearly see that it is set within a grid structure, with all the alignments and boxes neatly arrayed.

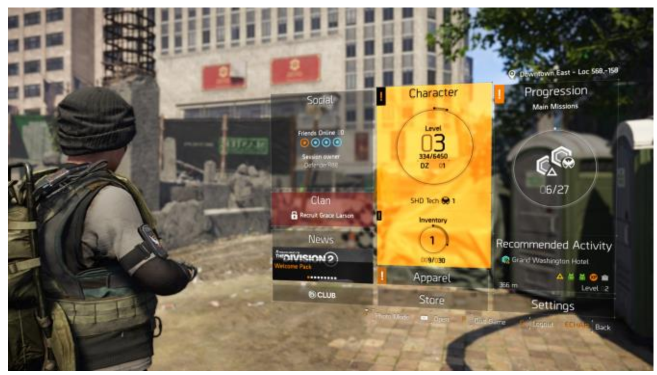

Figure 3. Example of grid elements in diegetic UI ("Tom Clancy's The Division 2", Ubisoft) [9].

B. *Hierarchy*

Hierarchy exists in all facets of games, from typography to the visual dominance of imagery. There is example of visual and typographic hierarchy observed in this UI for "Destiny 2" (Figure 4).

With value on simplicity, one sans serif typeface remains used (almost) exclusively throughout the game, with importance indicated by variances in weight, size/scale, case and spacing. In this example the headline and sub-heading are in all uppercase with alternate weights, whilst the body text is sentence or lowercase and much smaller. This typographic strategy produces a harmonious and clean result, which is easy to read and understand.

The large visual element of the 'Pariah' gun draws our attention to its features, thus also demonstrating visual hierarchy through placement and scale of image.

Figure 4. Typographic hierarchy (“Destiny2”, Bungie Games) [10].

C. Scale

Applying appropriate scale to UI is not only extremely important to prevent information from impeding the view, but also to create a UI which is legible and easily operated.

The examples from non-diegetic racing game, “Forza Horizon 5” (Playground Games) in Figure 5 and Figure 6, we demonstrate how effective use of scale through differentiation of image and text is utilised as a graphic device; to create visual hierarchy and draw attention dynamically to the information of most import. In Fig.5 the ‘1000’ text stands out as the dominant lead-in graphic, through its comparatively large scale on the screen. In Fig.6 the use of a large car image becomes the directive element, The eye looped back across the screen to the ‘Start Race Event’ box by its size and positioning.

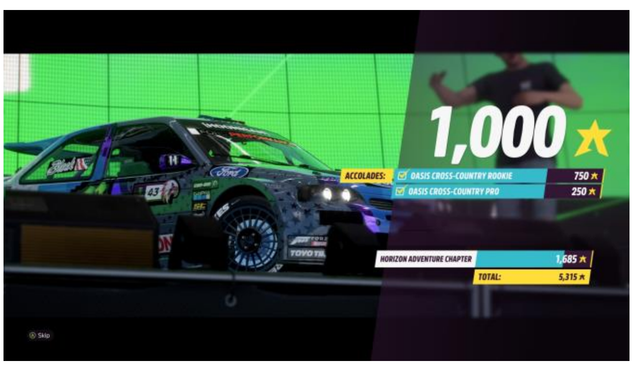

Figure 5. Example of scale (“Forza Horizon 5”, Playground Games) [11].

D. Balance (Symmetry and Asymmetry)

An isometric puzzler game, Monument Valley (Ustwo) is renowned for its zen-like qualities, benefitting from a sensitive use of colour. In this game, as seen in Figure 7 and Figure 8, the symmetry applied to each screen – with harmony achieved by the isometric structure – enhances this balance by use of a minimalist UI. The peaceful nature of the game visuals due in large part to the graphic design principles applied.

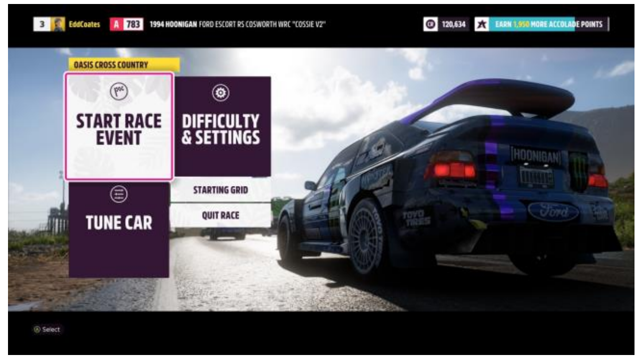

Figure 6. Example of scale (“Forza Horizon 5”, Playground Games) [12].

Figure 7. Symmetrical balance (“Monument Valley”, Ustwo) [13].

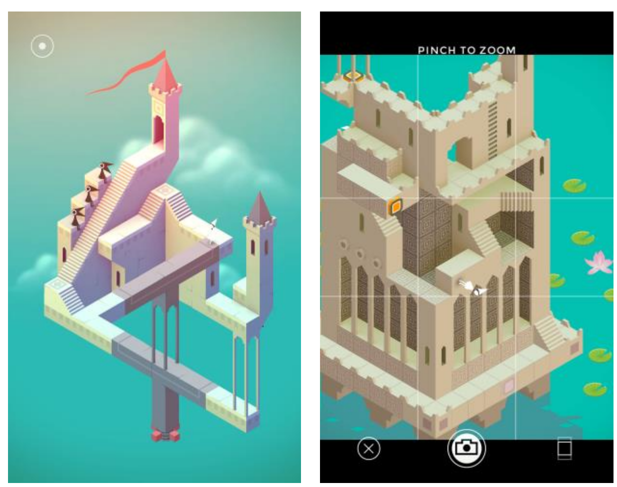

Figure 8. Example of symmetrical balance of visual and UI tools (“Monument Valley”, Ustwo) [14][15].

Similarly balanced but in a different genre, adventure game "Firewatch" (Campo Santo), as seen in Figure 9 and Figure 10, demonstrates a minimalist, light, and airy non-diegetic UI HUD with plenty of space surrounding text elements, centred text boxes for speech, and use of diegetic UI pieces throughout the gameplay, The elements appear to pivot in the middle of screen, as the player opens a map (Figure 10), or uses a compass, for example.

Figure 9. Symmetrical balance ("Firewatch", Campo Santo) [16].

Figure 10. Symmetrical balance ("Firewatch", Campo Santo) [17].

The perspective of the player is of foremost importance to the balance and symmetrical nature of the visuals here, and this has affected the design choices for a harmonious UI.

E. *Framing*

In this example of framing (Figure 11), from isometric dungeon crawler "Hades" (Supergiant Games), we see clever use of a dark, shadowy vignette to draw attention inwards from the edge of the screen, toward the speech box pop-up (also inset in a framed box). This device facilitates discussions with the Gods, however actual frames also appear as a regular element throughout the gameplay to hold information and communications (Figure 12) and direct the eye.

Figure 11. Example of framing ("Hades", Supergiant Games) [18].

Figure 12. Example of framing ("Hades", Supergiant Games) [19].

IV. THE ANALYSIS OF UNDERUTILIZED ELEMENTS OF GRAPHIC DESIGN IN COMPUTER GAMES

Aspects of graphic design in games are utilised well and often. Colour, for example, works with significant effect to make differentiations and highlight important touchpoints in most games. Gestalt Principles, and affordances, are also widely used and critical to how players both view the action and the UI elements, thus experiencing a game [1]. However, there are certain elements of graphic design which are either harder to incorporate onscreen or potentially overlooked.

A. *Dynamic Contrast / Asymmetry*

Few games include a truly dynamic UI. "Persona" (by P-Studio, Atlus) is a stand-out example of this, where the UI's highly graphic, contrast-heavy, and animated shapes flow with the player and give an extremely stylised, attention-seeking and rule-breaking [3] game experience (see Figure 13 and Figure 14). This game remains one of few to embrace a sense of order-in-chaos through asymmetry, contrast in colour and dynamic animation.

B. *Typographic Knowledge*

How to use typographic hierarchy effectively is a design principle which can be tricky on digital surfaces, and games are no exception. Selecting the correct and most appropriate typeface, for example, can have major influence on the result from a UI perspective. A font which does not "match" the character or 'feel' of the game, or is not clear enough in its communication, and legibility (particularly when sized down to UI level), can slightly lift the player out of a sense of immersion, or at worst actively inhibit the player from understanding what they need to do.

Figure 13. Asymmetry and contrast ("Persona" P-Studio, Atlus) [20].

Figure 14. Asymmetry and contrast ("Persona" P-Studio, Atlus) [21].

C. Simplicity & Clarity

Simplicity references the 'less is more' principle that most designers ascribe to – that simple is, in fact, hard to do well. Computer games, from in-game facets to UI and marketing material, trend toward a very narrative and/or illustrative style to the design work – which can be necessary, but also requires restraint in execution. With the importance of screen real-estate and the goal of player immersion HUD UI elements are typically simple and non-intrusive. With a simple and intuitive game design a game UI can in fact be almost non-existent. Games, such as Limbo and Inside (Playdead) uses no HUD and all UI elements are limited to small spatial tool tips to show interactivity.

V. Evaluation of Graphical User Interfaces in Games

Game GUI performs a vital role in the User Experience (UX). This basic analysis into graphic design principles examines whether these core elements present from a communication perspective – simply posing the question, "do the core design principles exist in this game?" Analysis of successful use is still subjective in nature – games with very aesthetic and fundamentally elegant UI may fail to address ergonomic or experience concerns from a technical perspective (do the UI elements work well in placement for console, for example, or does the UI work well across all screen types and sizes?). Addressing graphic concerns still go a long way to developing a great UI design. What we can determine from the application of the 5 core design principles in the analysis performed in this paper is all 5 graphic design principles are used to some extent in analysed games.

As a work of design, games are no different to any frame of reference which seeks to communicate using text and image. Successful visual communication occurs with the excellent execution of design basics. As per the conclusions drawn below, we believe that the games we examined all used the core graphic design principles (Section II), despite varying genre and platform.

1) Grid: All games analysed showed evidence of columns, boxes and alignment, and consistency of placement, which implies the use of a grid system. Some UI set-ups appeared to be more complex than others, requiring a grid which was at once structured and flexible. This speaks to the genre of each game. For example, where "Monument Valley" (designed as a mobile app) is by necessity very simple, touchscreen orientated and reliant on user intuition rather than an expansive UI, "Destiny 2" (designed for PC and console) has a long narrative arc, involving many and varied elements (such as a curved diegetic HUD hinting that its projected on the players helmet visor and non-diagetic UI screens, icons and tools and meta pop-up layers).

2) Hierarchy: All games analysed used some form of typographic element, with associated hierarchial features, to pass on information to the player - be they complex as 'Destiny 2' or 'Forza Horizon 5', or more simple in their communications like 'Firewatch' or 'Monument Valley'. Visually, hierarchy was also achieved through use of colour, layering and transparency, to make certain UI features stand out as the player required.

3) Scale: Knowing appropriate scale to use for digital outputs is crucial to a working UI. If button scale is too large, or text too small, the usability and balance of design is thrown out. All the games analysed in this paper displayed variance, but addressed issues of scale appropriately. Some used scale to bring hierarchy to their UI designs whilst others were more dependent on visual contrasts and scale to lead the player.

4) Balance: Some of our analysed games, such as 'Firewatch' and 'Monument Valley' were clearly quite symmetrical in structure, with centred elements and space around the action onscreen. This applied to their UI as well, which appeared simple and easy to identify. Deliberate asymmetry was harder to find, beyond stand-out examples, such as "Persona 5" which straddled the line between chaos and order. Asymmetry for balance in all games was present to a small degree, but in terms of the UI elements edged much closer to favouring symmetrical and harmonious, easy to navigate results.

5) Framing: All games analysed adapt to a frame of some kind (and have a margin to work within, where no information will cross). Some, like "Hades", use the more literal version of the concept to put containers around UI elements. All of the games analysed used boxes for information and/or imagery in their UI executions at some point, and many (like "Forza Horizon 5") dynamically cropped imagery in their UI screens to help lead the eye.

VI. CONCLUSION

Graphic design principles, applied appropriately and with consistency and forethought, can increase positive user experiences with game UI and assist in the clear communication of information critical to gameplay. In an analysis of 5 core design principles as they apply to a range of computer games, there is evidence of all being utilised, despite the disparity in game genre (be it isometric, first-person shooter, action-adventure, racing or Zen puzzler). The complexity of game structure and choice of UI style often dictates to what extent the graphic design principles play a role. This paper presents an analysis into graphic design principles to examine whether the core elements present from a communication perspective. In future studies, we plan to conduct a set of user studies to measure user experience and test the players' immersion in gameplay to investigate whether the UI elements work well in placement for console and across all screen types and sizes.

REFERENCES

[1] E. Lupton, "Design is storytelling", Cooper Hewitt, New York, The Gaze, pp.118-125, Gestalt Principles, pp.126-131, 2017.

[2] E. Lupton and J. Phillips, "Graphic design: the new basics" [ebook] San Francisco, Chronicle Books, pp.8-11, pp.60-67, pp.116-139, pp.186-199, 2015.

[3] A. Dakinedi, "Top 5 Best Video Game UIs", Available from: https://superjumpmagazine.com/top-5-best-video-game-uis-db941d6a9357, 2018, retrieved: March 2022.

[4] Sourcegaming, "Holism: Metroid Prime and the Scan Visor", https://sourcegaming.info/2017/06/28/holism-metroid-prime-and-the-scan-visor/, 2017, retrieved: March 2022.

[5] A. Hayward-Cole, "7 basic principles of graphic design", Available from: https://www.lcca.org.uk/blog/careers/7-basic-principles-of-graphic-design/, 2019, retrieved: March 2022.

[6] E. Lupton and J. Phillips, "Graphic design basics: core principles for visual design", Skillshare, Available from: www.skillshare.com/classes/Graphic-Design-Basics-Core-Principles-for-Visual-Design/15397, 2021, retrieved: March 2022.

[7] M. Bowers, "Level up: a guide to game UI (with infographic)", Available from: https://www.toptal.com/designers/gui/game-ui, 2020, retrieved: March 2022.

[8] Interfaceing Game, Destiny 2, "destiny-2-customize-face.jpg", Available from: https://interfaceingame.com/screenshots/destiny-2-customize-face/, 2017, retrieved: March 2022.

[9] Interfaceing Game, Tom Clancy's The Division 2, "tom-clancys-the-division-2-character-profil.jpg", Available from: https://interfaceingame.com/screenshots/tom-clancys-the-division-2-character-profil/, 2019, retrieved: March 2022.

[10] Interfaceing Game, Destiny 2, "destiny-2-rifle.jpg", Available from:https://interfaceingame.com/screenshots/destiny-2-rifle/, 2017, retrieved: March 2022.

[11] Game UI Database, Forza Horizon 5 'Rewards and Experience, "Forza-Horizon-511112021-124736-95852.jpg", Available from: https://www.gameuidatabase.com/gameData.php?id=1225#&gid=1&pid=45, 2021, retrieved: March 2022.

[12] Game UI Database, Forza Horizon 5 'Pre-Game & Lobby', "Forza-Horizon-511112021-124734-38594.jpg", Available from:https://www.gameuidatabase.com/gameData.php?id=1225#&gid=1&pid=39, 2021, retrieved: March 2022.

[13] Monument Valley Wiki, "Geometric-Building-Levels-in-Monument-Valley-Game-by-ustwo.jpg", Edited, Available from:https://monumentvalley.fandom.com/wiki/The_Spire?file=Geometric-Building-Levels-in-Monument-Valley-Game-by-ustwo.jpg, 2019, retrieved: March 2022.

[14] Moby Games, Monument Valley Screenshots, "781657-monument-valley-android-screenshot-many-of-the-puzzles-are.jpg" Available from: https://www.mobygames.com/game/android/monument-valley/screenshots/gameShotId,781657/, 2015, retrieved: March 2022.

[15] Moby Games, Monument Valley Screenshots, "781648-monument-valley-android-screenshot-camera-mode-while-ida-is.jpg" Available at: https://www.mobygames.com/game/android/monument-valley/screenshots/gameShotId,781648/, 2015, retrieved: March 2022.

[16] Game UI Database, Firewatch 'Title Screen', "Firewatch07102020-111051-72794.jpg", Available from: https://www.gameuidatabase.com/gameData.php?id=238#&gid=1&pid=1, 2016, retrieved: March 2022.

[17] Game UI Database, Firewatch 'Area Map', "Firewatch07102020-111055-51693.jpg", Available from: https://www.gameuidatabase.com/gameData.php?id=238#&gid=1&pid=9, 2016, retrieved: March 2022.

[18] Interfaceing Game, Hades 'Dialogue', "hades-dialogue.jpg", Available at https://interfaceingame.com/screenshots/hades-dialogue/, 2018, retrieved: March 2022.

[19] Interfaceing Game, Hades 'Well of Charon', "hades-well-of-charon.jpg", Available from: https://interfaceingame.com/screenshots/hades-well-of-charon/, 2018, retrieved: March 2022.

[20] Game UI Database, Persona 5 'Player Menus', "Persona-511142020-081939-92444.jpg", Available from: https://www.gameuidatabase.com/gameData.php?id=72#&gid=1&pid=46, 2016, retrieved: March 2022.

[21] Game UI Database, Persona 5 'Modal: Item Get', "Persona504202020-043344.jpg", Available from: https://www.gameuidatabase.com/gameData.php?id=72#&gid=1&pid=25, 2016, retrieved: March 2022.

[22] K.Lohse, N. Shirzad, A.Verster, N. Hodges, and H.M.Van der Loos "Video games and rehabilitation: using design principles to enhance engagement in physical therapy" Journal of Neurologic Physical Therapy, 37(4), 166-175, 2013.

[23] W. Cai and V.C. Leung, "Decomposed cloud games: Design principles and challenges". In 2014 IEEE International Conference on Multimedia and Expo Workshops (ICMEW) pp. 1-4. IEEE. 2014, July.

Deep Learning-Based Food Identification and Calorie Estimation System

Jung-Tang Huang
Taipei Tech, 1, Sec 3, Chung Hsiao E, Road, Taipei 106,
UTL Lab
Taiwan, Taipei
jthuang@ntut.edu.tw

Chen-Hao Wang
Taipei Tech, 1, Sec 3, Chung Hsiao E, Road, Taipei 106,
UTL Lab
Taiwan, Taipei
kevin8901300179@gmail.com

***Abstract*—In recent years, the importance of home care and diet have increased due to an aging population and increasing care needs. This study proposes a solution to take a photo every time while the user is eating, record the photo and performing identification and calorie estimation to understand the diet throughout the day. This paper uses a depth camera for image recognition and calorie estimation of food, and uses more than 400 food photos as a database to train a YOLOv5 neural network to recognize food on a table. The depth camera is then used to estimate how many calories the user has consumed. Our experimental results show that a food recognition rate of more than 90% can be obtained in most cases, while the calorie estimation error is less than 10%. Then, we use Google's Firebase service to upload and store a large amount of data and build a complete Internet of Things (IoT) care system with these large amounts of dietary data. This contributes to a complete elderly home care.**

Keywords- Image recognition; Volume estimate; Calorie estimates; Depth camera; Deep learning.

I. INTRODUCTION

In recent years, sarcopenia and obesity in the elderly have been global health problems with broad economic and social implications. Nutritional monitoring systems are essential for understanding and addressing unbalanced eating habits. Therefore, in order to maintain a healthy weight in a normal person with a healthy diet, daily food intake must be measured. Daily food intake and nutrient distribution are particularly important, which requires patients to calculate and record calories from food every day. In most cases, patients also have a lot of trouble estimating food intake due to self-denial issues, lack of nutritional information, and the time-consuming process of manually recording this information. The most common way to measure nutrients in food today is to directly use the weighing method. Whether it is taking out the dishes from the plate after cooking the food or measuring the food before it is cooked, it takes a lot of time. In addition, it is a big trouble to measure all the data and record it using paper or electronic products, so a system developed in this paper covers calorie estimation and cloud service system.

The rest of this paper is organized as follows: Section 2 gives the literature review. Section 3 provides an overview of data-related work. Section 4 describes calorie content estimates. Section 5 discusses the results. Section 6 presents the conclusions and suggests possible future directions.

II. RELATED WORKS

A. *Food identification related literature*

Recognizing food in images is a difficult object detection task. There are many machine learning methods to detect objects. Some researchers use Histogram of Oriented Gradients (HOG) features [1] and Scale-Invariant Feature Transform (SIFT). Another approach that has been increasingly used in recent years is deep learning neural networks, especially Convolutional Neural Network (CNNs). CNNs work very well and are probably the most widely used object detection method in the world. Some progress has been made in this area, and more object detection techniques have been designed based on CNNs. Some well-known object detection CNN designs are Mask R-CNN (Mask Region Convolution Neural Network) [2], Faster RCNN [3] and YOLO [4]. Most of the previous studies on image-based food calorie estimation have used detection classification to obtain food category labels. Deep learning neural networks have recently gained considerable application in object detection. Wearable context-aware food recognition for calorie monitoring [5], and segmentation and recognition of multi-food meal images for carbohydrate counting [6], both employ food segmentation-based approaches to address issues related to diet assessment. In [7], T. Ege and K. Yanai developed an object detection model using Faster R-CNN. The authors also propose a calorie estimation procedure in a multi-task CNN [8] to estimate food categories and calories simultaneously. Their model achieved high average accuracy detection (90.7%), but the paper did not use a weight estimation procedure, so the estimated calories for each food were fixed.

B. *Food estimation literature*

In [9], the authors use Faster R-CNN for object detection by using top and side views of food. They also used the GrabCut algorithm [10] to obtain the contours of each food. Finally, food size estimation and calorie estimation use estimation formulas done using volume. The only downside to this method is that it only works with single-item images of food, such as fruits, vegetables, and so on. In [11], the authors use CNN-based multi-task learning for dietary assessment of multiple food dishes. They achieved high accuracy in food segmentation and volume estimation. The authors of [12] completed a volume-based approach to food calorie estimation using depth cameras. Their approach

 ISBN: 978-1-61208-978-2

emphasizes the use of in-depth information to predict the volume estimation part of food calories. In [13], K. Ruan and L. Shao used Support Vector Machines (SVM) and deep learning algorithms for food classification. For the calorie estimation section, they created a calorie map for all food labels.

III. DATA ACQUISITION

When training deep models, increasing the amount of data as much as possible is desirable; the more the better. Therefore, when collecting data, the same object should be looked at from different angles, from different distances, and even illuminated with different lights. One should shoot with a body camera or scope, or try different aperture and shutter variations. In this way, the trained model can be used in various environmental fields. In this work, a food dataset named Food-101 [14] is used, which contains a large number of images. The dataset contains 101 categories and more than 10,000 photos. We choose home-cooked dishes often eaten by oriental people for discussion. This article uses more than 200 photos in 4 categories in the data set, plus the pictures we collected; the total number of photos is more than 400 photos, and the categories include: rice, eggs, shrimp, broccoli, etc.

IV. METHODOLOGY

A. YOLO

The full name of YOLO is ‘you only look once’, which means that you only need to browse once to identify the category and location of the object in the picture. Because it only needs to be seen once, YOLO is called the Region-free method. Compared with the Region-based method, YOLO does not need to find the possible target, Region, in advance. That is to say, the process of a typical Region-base method is as follows: first, analyze the picture by means of computer graphics (or deep learning), find out several areas where objects may exist, cut out these areas, put into an image classifier, and classified by the classifier.

B. Pixel and Length Mathematical Model

Before using the depth camera to calculate the real volume, one must first understand the real length and width of the object in the depth image, so there is a need to calculate the length of each pixel in the depth image. The relationship equation between the number of pixels of the object and the length and width of the object needs to be obtained through the following methods:

(1) The depth camera shoots an object with exact length.
(2) The number of pixels corresponding to the object in the image when the depth camera object is taken from far to near.
(3) After obtaining the corresponding pixel amounts at different depths, an equation is obtained.

The relationship between pixel and length obtained in the experiment is recorded in Table 1. We tested seven different depth distances, which corresponded to seven different pixel amounts.

TABLE I. PIXEL AND LENGTH RELATIONSHIP

Depth (cm)	Number of Pixels (Pixels)	The length corresponding to the pixel (cm/pixel)
92	101	0.1509
90	103	0.1480
87	107	0.1425
93	112	0.1361
74	125	0.1220
68	137	0.1110
43	219	0.0696

The relationship between pixels and length can be deduced from the linear equation (1), from which we can know the length of each pixel of the camera at any depth value:

$$Lenght = 0.0016618 * Depth + 0.0017478 \quad (1)$$

The actual area of the pixel is to square the linear equation (1), and the equation is as follows (2):

$$Area = (0.0016618 * Depth + 0.0017478)^2 \quad (2)$$

C. Volume estimation

Calculate the depth difference of each pixel coordinate, and express the depth difference as $\Delta\mathrm{Depth}$. Mark the depth value before the meal as Depth_0, and mark the depth value after the meal as Depth_1, calculate the depth difference of the pixel point:

$$\Delta Depth = Depth_0 - Depth_1 \quad (3)$$

The volume of a pixel is the area of the pixel multiplied by the depth difference of the pixel, where k is the coordinate of the pixel:

$$V_k = \Delta Depth_k * Area_k \quad (4)$$

Finally, add up all the pixel volumes to get the volume of the food.

D. Weight Estimate and Calorie Estimation

When the volume of the food is calculated, the next step is to convert the volume to weight and the method we use is to use the density formula for conversion, where m is the final weight of the food, v is the estimated food volume, ρ is the food density:

$$m_{food} = \rho_{food} * v_{food} \quad (6)$$

Finally, convert the food calorie dataset corresponding to the calculated food weight and food category into calories, and three major nutrients.

E. Cloud Service

Cloud Storage for Firebase stores files in Google Cloud Storage buckets, thus, they can be accessed through Firebase and Google Cloud. Flexibility through the Firebase SDK for Cloud Storage uploads and downloads files from mobile clients. Additionally, it is possible to use the Google Cloud Storage API Server-side processing, such as image filtering or video transcoding. Cloud storage scales automatically, which means no need to migrate to any other provider.

We collate the results of identification and calorie estimates and upload them to Firebase, which can increase the amount of data. In the future, due to the increase in the amount of data, the number of photos during training can be increased, the recognition rate can be improved, and the number of categories can even be increased. This will help with the user's eating habits and making dietary recommendations for the user.

F. *Hardware*

1) *Experimental Environment*

In the experimental environment part, we use Intel D435i depth camera and Raspberry Pi 4B. The camera shoots from top to bottom to obtain a deep image of the plate, and the distance between the camera and the platform is 60 cm.

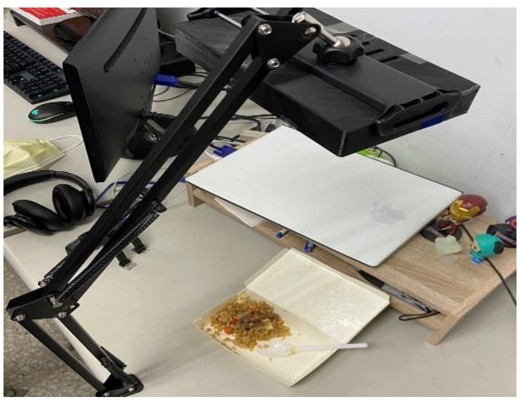

Figure 1. System experimental environment

2) *Raspberry Pi 4B*

In recent years, the topic of IoT has been discussed vigorously, and many developers have used Raspberry Pi to develop various IoT devices. In this study, Raspberry Pi 4B was used as the development board of this research as the router of the system, the development program and the calculation of calorie estimation program and the center of uploading the values of various sensors to the cloud.

3) *Intel D435i*

The Intel RealSense Depth Camera D435i incorporates an Inertial Measurement Unit (IMU) that improves depth perception in any situation where the camera moves. The D435i camera can be used in SLAM and tracking, and can also improve the efficiency of point cloud computing. It can be used on different platforms such as computers or embedded development boards. This study uses D435i as our camera, which can be used to identify and detect the depth of the food, and finally get accurate data.

V. RESULTS

A. *Image recognition results*

In this section, we train YOLO's food recognition model with 1000 steps. Figure 2 shows the loss function used to train the YOLO model. As for the identification results, we choose home-cooked dishes that Orientals often eat for discussion, randomly select 10 images from the dataset, and use our trained model for identification. The average accuracy rate for individual food items is over 90%. The recognition results are shown in Figure 3, and the recognition rate is above 93%.

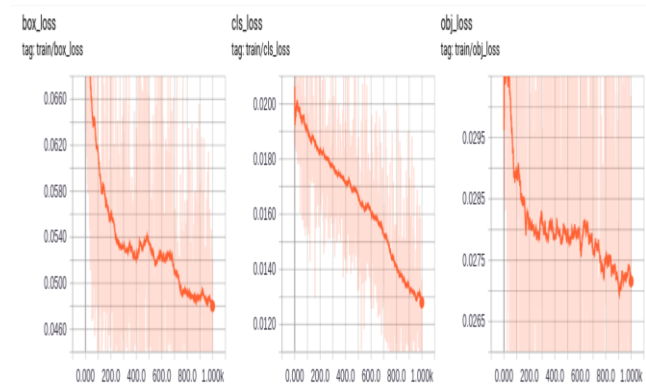

Figure 2. Loss function for YOLO model

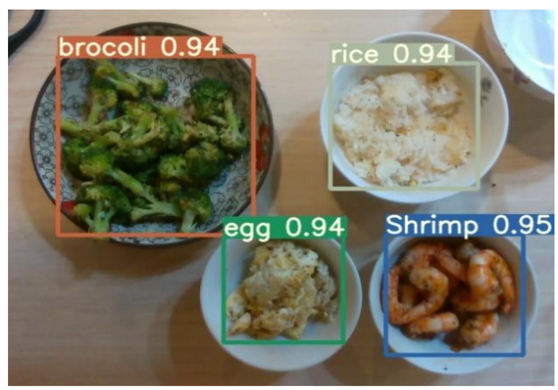

Figure 3. Image recognition results and recognition rates

B. *Calorie Estimation Results*

This result is the same as the previous experiment. We randomly select 10 images from the dataset and use our model to estimate the calories of the food, record the real weight and the estimated weight, average all the recorded weights and calculate the error. Table 2 shows that the calorie estimate has an error of less than 10%, which means that the calorie error will also be less than 10 calories.

 ISBN: 978-1-61208-978-2

TABLE II. CALORIE ESTIMATION TEST RESULTS

Food	Data		
	Actual weight (g)	Estimated weight (g)	Error (%)
Rice	82.8	76.8	7.8
Egg	71	72.8	2.5
Broccoli	70.6	65.8	7.2
Shrimp	97.6	103.8	6.3

C. *Cloud Service*

In this section, we upload the results of the diet identification and calorie estimation to the cloud and classify them. The type and portion of each food are listed and clearly marked, as shown in Figure 4.

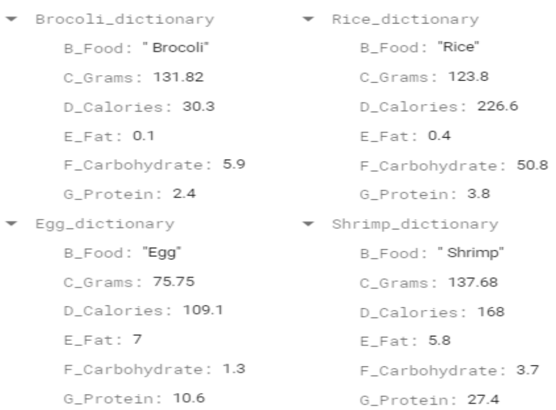

Figure 4. Firebase Profile

VI. CONCLUSION

The results show that the meal estimation method using the depth camera proposed in this paper is feasible for food evaluation, but the premise is that all the food in the plate can use the depth camera to obtain accurate depth values.

This method is compared with the traditional weighing method. It is very convenient to measure food items one by one, which is convenient for caregivers in nursing homes who spend a lot of time measuring and recording the dining conditions of the elderly.

In this article, we use YOLO to detect food calories for different foods. The recognition rate is over 90%, and the calorie error is less than 10%. However, estimating food calories is a difficult problem. Not even a perfect image processing system can predict perfectly, and the calorie content can be skewed by the amount of oil. Therefore, we first try to take the prototype food as the research direction. In the future, we will increase the diversity of oily and water-based fried vegetables and tofu, so that the dietary information can be easily obtained by using the meal evaluation method of this study in daily life.

Uploading data to Firebase in the future can increase the amount of data. In the future, due to the increase in the amount of data, the number of photos during training can be increased, the recognition rate can be improved, and the number of categories can even be increased. The user's eating habits and dietary recommendations for the user.

REFERENCES

[1] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), pp. 20–25 June 2005.

[2] H. Kaiming, G. Georgia, D. Piotr, and G. Ross, “Mask R-CNN,” Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 2961–2969, 2017.

[3] R. Shaoqing, H. Kaiming, G. Ross, and S. Jian, “Faster R-CNN: towards real-time object detection with region proposal networks,” NIPS 15: Proceedings of the 28th International Conference on Neural Information Processing Systems, Vol. 1, pp. 91–99, Dec 2015.

[4] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified real-time object detection", Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779–788, 2016.

[5] G. Shroff, A Smailagic, and D. P. Siewiorek, “Wearable context-aware food recognition for calorie monitoring,” 2008 12th IEEE International Symposium on Wearable Computers, pp. 1–2 Oct, 2008.

[6] M. Anthimopoulos ,J Dehais, and P Diem, “Segmentation and recognition of multi-food meal images for carbohydrate counting,” in 13th IEEE International Conference on BioInformatics and BioEngineering, pp. 10–13, Nov. 2013.

[7] T. Ege and K. Yanai “Estimating Food Calories for Multiple-Dish Food Photos,” 2017 4th IAPR Asian Conference on Pattern Recognition (ACPR), pp. 309–314, Nov. 2017.

[8] A. H. Abdulnabi, G. Wang, J .Lu, and K. Jia, “Multi-Task CNN Model for Attribute Prediction,” in IEEE Transactions on Multimedia, vol. 17, pp. 1949–1959, Nov. 2015.

[9] Y. Liang and J. Li, “Deep Learning-Based Food Calorie Estimation Method in Dietary Assessment,” Cornell University, pp. 12, Jun 2017.

[10] C. Rother, V. Kolmogorov, and A. Blake, “GrabCut: Interactive foreground extraction using iterated graph cuts,” ACM Trans. Graph., vol. 23, pp. 309–314, 2004.

[11] Y. Lu, D. Allegra, M. Antonopoulos , and F. Stanco, “A multi-task learning approach for meal assessment,” CEA/MADiMa '18: Proceedings of the Joint Workshop on Multimedia for Cooking and Eating Activities and Multimedia Assisted Dietary Management , pp. 46–52, July 2018.

[12] Y. Ando, T. Ege, J. Cho, and K. Yanai, “DepthCalorieCam: A Mobile Application for Volume-Based FoodCalorie Estimation using Depth Cameras,” MADiMa '19: Proceedings of the 5th International Workshop on Multimedia Assisted Dietary Management, pp. 76–81, Oct 2019.

[13] H. Raikwar, H. Jain, and A. Baghel, “Calorie Estimation from Fast Food Images Using Support Vector Machine,” International Journal on Future Revolution in Computer Science & Communication Engineering, vol. 4, pp. 98–102, April 2018.

[14] L. Bossard, M. Guillaumin, and L. VanGool, “Food-101– Mining Discriminative Components with Random Forests,” European Conference on Computer Vision ECCV 2014, pp. 446–461, 2014.

 ISBN: 978-1-61208-978-2