Linking Computerized and Perceived Attributes of Visual Complexity

Kanaka Babshet

School of Electrical and
Information Engineering
University of the Witwatersrand
Johannesburg, South Africa
Email: kanaka.babshet@fnb.co.za

Vered Aharonson

School of Electrical and
Information Engineering
University of the Witwatersrand
Johannesburg, South Africa
Email: vered.aharonson@wits.ac.za

***Abstract*—Psychological studies explore visual complexity as perceived by humans. Image complexity is studied extensively in the mathematical, computational sciences. The two disciplines often define visual complexity differently and are thus disjointed. This is manifested in differences between subjective human-perceived complexity, and computer vision algorithms' performance in visual tasks. Our study investigates this discrepancy in the context of cognitive tests that employ visual stimuli to assess a subject's primal cognitive functions. A database of cognitive tests including visual recognition tasks and the performance of 403 subjects in terms of response times was used. Computer vision and information theory features were extracted from the images in these tasks. Inspired by cognitive psychology studies, the features were categorized into whole-image and object-specific features. A random forest classifier was used to map the computed features into three complexity labels in the tasks, labelled according to the subjects' performance. The classifier computationally captured the significant features for the human-perceived task complexity by mapping the occurrence of these features to the complexity labels of the subjects' performance. The whole-image features demonstrated greater visual significance than the object-specific features. The features' importance values could provide insights into the links between mathematical visual complexity definitions and visual complexity as perceived by humans.**

Keywords– Visual complexity; Cognitive assessments; Computer vision; Binary images.

I. INTRODUCTION

Human visual perception is the processing and interpreting of a visual environment, transmitted from the eyes via neural paths to the brain. The image properties extracted in this activity culminate in a decision or action [1]. The details of the translation or encoding entailed in this process are, however, unknown.

Cognitive tests are designed to challenge this brain process. These tests display visual stimuli, pose a task associated with these stimuli, and require a response or decision from the tested subject. The performance of cognitively intact individuals in these tests could thus provide insights into the characteristics of the process involved in visual perception. Specifically, the differences between different tasks in terms of complexity could be assessed.

Understanding task complexity allows us to better engineer the interface of these cognitive tests. Having a predefined complexity scale offers a platform for dynamic adaptation to a user's cognitive capabilities by adjusting the complexity of the set of tasks presented based on their previous task response times. This ensures that the subject is presented with a task of an appropriate difficulty level for them, rather than something that is too easy or too difficult, which could cause frustration or boredom, and limit the usefulness of the assessment. An evaluation of performance which takes into account response times can also be refined by considering the complexity of the task as a weighting component in the test score.

Visual perception and the complexity of images were studied in cognitive psychology [2]. Witkin et al. [3] proposed a field dependence concept, explaining how people assess their visual field by either separating and organising the visual information into clear-cut groupings, or assessing their visual field as a whole. Attneave et al. [4] studied the different ways in which the perceived visual complexity of an image is affected by the information distribution in the image. Equivalently, computer vision studies employed mathematical image processing to extract information from an image [5]. The complexity of visual stimuli or images can thus be computed using mathematical metrics and computer algorithms.

Both disciplines share similar concepts conceptually on visual attributes on the information of a full image and/or part of image. Both strive to define visual complexity. The relevance of computer features and metrics to the way humans perceived complexity according to the aforementioned psychological theories is, however, rarely assessed.

Computerised cognitive tests are a context where visual stimuli in the form of computer generated images and a large cohort of human subjects performance can be studied. This study employs computerised cognitive tests data and aims to find a set of computed attributes, or features, which could help explain the complexity of a task associated with visual stimuli.

This paper first presents a short background on the test data provided for this study, with the details of the subsequent algorithm development and implementation in Section 2. It is then followed by the feature results in Section 3, and a concluding analysis and discussion in Section 4.

II. METHODS

Test data from previously conducted cognitive tasks was made available for this study. This was used to find a set of computed visual attributes which could correlate to the tasks' perceived complexity based on the task results in the provided data. This Section presents the details of this process.

A. The Data: Visual Stimuli and Human Performance Data

The cognitive tasks and users' performance data were taken from NexSig's computerised cognitive testing studies [6]. The visual part of this dataset involves stimuli in the form of simple, black and white, four-by-four square images that can

 ISBN: 978-1-61208-780-1

be described as 16-bit binary arrays, as illustrated in Figure 1. In this study, we focused on one type of visual task from this test battery: a recognition task. An example recognition task is illustrated in Figure 1: three images are presented on the screen, and the subject is required to recognise which of the three images is different.

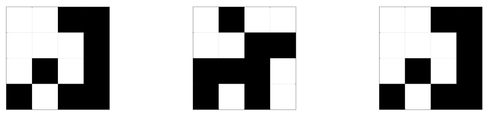


Figure 1. An example of a **recognition** task.

The computerized tests recorded and logged both response time and correctness of the subjects' responses for each of the tasks presented. All subjects in the studies were assessed by healthcare professionals prior to taking the cognitive tests. The database used in our study included tests of cognitively intact subjects only, who had no motor or vision impairment.

There were 5087 recognition tasks in the dataset. The data was stored in a table, where each row represented a task instance. The images presented in the tasks were coded as 16 bit arrays where a white square in the image is represented by a 0 and a black square by a 1. A binary array encodes the image starting at the top left and downwards row-by-row. For example, the first image in Figure 1 would be encoded as 0011000101011011. Following the 3 binary vectors of the 3 images presented in the task, each row contains the subject's response and response time in milliseconds. A separate table contains the demographics of the subjects: age, gender and computer proficiency.

B. Algorithm Implementation

Figure 2 depicts a flow diagram of the methodology employed in the study.

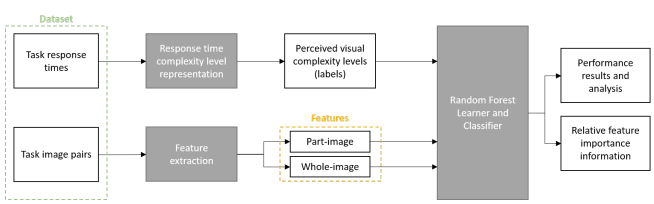


Figure 2. A flow diagram of the algorithm implementation

Each response time in the dataset was represented by a label corresponding to a human-perceived complexity. Concurrently, visual features of the cognitive tasks were extracted. The human perceived complexity labels and the features were the inputs to a random forest learner. The machine-learned selection process indicated which of the features were relevant in the human-perceived visual complexity prediction. The implementations of the blocks in Figure 2 are described below.

1) Human Visual Complexity Level Representation: An inherent assumption in this study, corroborated by the administrating neuropsychologists, was that the subjects' response times were associated with, or reflected in, the tasks difficulty level or complexity. Subject demographics - age, gender and computer proficiency - were examined to ensure that these factors did not distort the response-times distribution. Initial examination of the dataset response times yielded that their distributions were similar, and had a Gaussian shape with a longer right-hand tail, for all age groups, for male and female subjects and for groups of proficiency levels. The response-times' continuum was segmented into K segments from which a complexity labels scale was constructed, from "easy" to "hard," corresponding to the segments of short response times to the segments of long response times, respectively. Different segmentation paradigms, as well as different K values were applied and evaluated in different experiments, as labels for the random forest learner.

2) Computer-Vision Complexity Attributes: The algorithm was developed to discover which visual features could define the visual complexity in a recognition task. The study was performed on the pair of the 2 different images from the 3 images presented in the task. This choice was based on the assumption that the third image, identical to one of the images in the pair, does not significantly impact on the visual comparison.

3) Feature Extraction from the Image Pair: The features were either adapted from earlier image processing and computer vision studies, or were conceptually based on Attneave [4] and Witkin's [3] psychological theories on visual perception.

a) Object Type Definitions: The following visual object *types* are referred to in the subsequent feature descriptions. All object types are defined for both black and white blocks:

- **Adjacent Path:** Consecutive adjoining blocks of the same colour (directly next to each other, against one of the four **sides**) to form a path of its own. Black adjacent paths outlined in red in the example image (Figure 3).
- **Diagonal Path:** Consecutive blocks of the same colour diagonal to each other (against one of the four **corners**) to form a path of its own. Black diagonal path outlined in green in the example image (Figure 3).
- **Single block:** a block that is not part of an adjacent path or diagonal path. Black single block outlined in blue in the example image (Figure 3).

By these definitions, the image in Figure 3 has a white adjacent path, and a white single block as well.

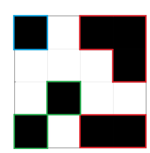


Figure 3. Example image to illustrate the object **types**.

b) Feature Extraction Paradigm: Quantifying a recognition task's complexity entails a consideration of a *relative* visual complexity of a pair of images, i.e., if one image is visually complex, while the other is simple, the recognition task will be relatively easy.

All features were computed for each image in the pair, and inserted into the machine learning model as two independent features. Additionally, features that pertain to a comparative nature were computed for the pair of the images and were used as a single feature. The latter, relative type of features are marked in the list below with a "(relative)" next to the feature name.

Inspired by the cognitive theories of Attneave [4] and Witkin [3], two categories of features were calculated: object-specific features, and whole-image features. Examples and illustrations for these features are described further.

c) Object-Specific Features:

- **Number of objects**: This feature is a count of the number of adjacent paths, diagonal paths, and single blocks respectively within each image, for both black and white objects. Table I presents an example of this count, computed for the image pair presented in Figure 4.

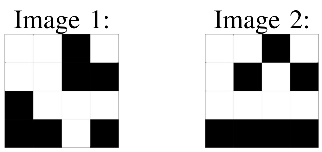


Figure 4. Example image pair to illustrate *number of objects*, *object path lengths*, *different objects present*, and *similar objects within an image*

TABLE I. NUMBER OF OBJECTS FOR THE IMAGE PAIR IN FIGURE 4

	Image 1		Image 2	
	Black	White	Black	White
Adjacent Paths	2	1	1	1
Diagonal Paths	0	0	1	0
Single Blocks	1	1	0	1

- **Object path lengths**: This feature is the total lengths of the adjacent paths and the diagonal paths, for both black and white objects, in an image. Table II presents an example for this feature, computed for the images of Figure 4.

TABLE II. OBJECT PATH LENGTHS FOR THE IMAGE PAIR IN FIGURE 4

	Object Path Lengths			
	Image 1		Image 2	
	Black	White	Black	White
Adjacent Paths	6	8	4	8
Diagonal Paths	0	0	3	0

- **Objects with similar angles (relative)**: This is the number of black or white objects that have either the same angle, or the inverse angle in a pair of two images, measured from the horizontal axis.
- **Objects with similar locations (relative)**: This feature indicates the number of objects that have **similar** locations. The locations are calculated as a centroid midpoint and points within half a square of each other are considered as similarly located objects. Only objects of identical type and colour, i.e., black adjacent paths, are checked for location similarity. Figure 5 illustrates similar locations for a pair black adjacent paths, a pair of white adjacent paths, and a pair of white single blocks in the two images.

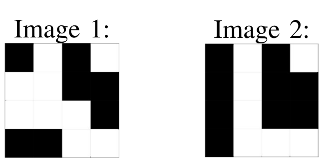


Figure 5. Example image pair to illustrate *objects with similar locations*

- **Different object types present (relative)**: This binary feature returns a one if any of the previous object types exists in one image, but not in the other. For example in Figure 4, Image 2 has a black diagonal path while Image 1 does not, while Image 1 has a black single block while Image 2 does not, which will produce a value of one.
- **Similar objects within an image**: This feature indicates the number of objects that have similar shape and size, regardless of angle, within an image. In Figure 4, for example, image 1 has two similar objects, while Image 2 has none.
- **Similar objects in an image pair (relative)**: This feature counts the number of objects that have similar shape and size, regardless of angle, in both images of the pair. For example, there is one similar object in the two images of Figure 6.

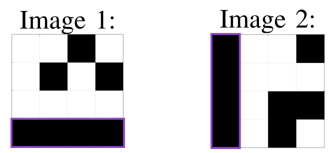


Figure 6. Example illustrating *similar objects found in an image pair*

d) Whole-Image Features:

- **Whole-image object spacing**: This feature is an average of the distances between objects of one colour in an image. In the example of Figure 7, the average distance for the black objects in Image 2 is larger than in Image 1.

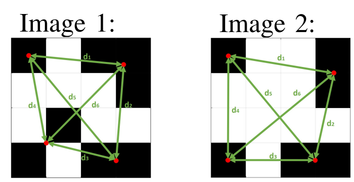


Figure 7. Example image pair showing the various object distances calculated for *whole-image object spacing*

- **Whole-image squares comparison (relative)**: This feature is a direct comparison of an image pair, where each square in the image is referred to as a bit; black represented by 1 and white by 0. The computation entails an XOR on the 16-bit array representations of the images and a summation of the resulting array.
- **Relaxed image symmetry**: This is a binary feature that indicates symmetry, defined across the horizontal axis, the vertical axis, and the two diagonals, including inverse colour symmetry as illustrated in Figure 8.

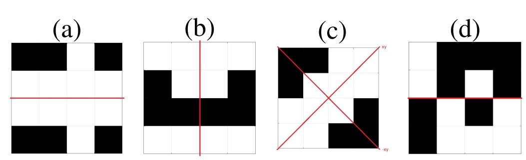


Figure 8. Example images which display the various definitions of symmetry

'Relaxed' symmetry, defined as one-block difference within the symmetrical image, is also considered as an

occurrence of symmetry due to the low prevalence of exact symmetry in these images.
- **Features from Gabor filters**: Gabor filters are made up by sinusoidal planes modulated by a Gaussian envelope [7]. A ‘filter bank’ of Gabor filters models the first stage of the brain’s visual processing (V1) [8]. In our implementation, a filter bank of Gabor filters at various orientations and frequencies was applied to each image in the pair, resulting in an output feature vector for each image. The features were the sum and standard deviation of the vector.
- **Fractal dimension features**: The ratio that indicates the level of visual detail in a fractal pattern at different levels of magnification was calculated using the box counting technique [9]. This technique estimates the number of boxes required to cover the non-zero parts of an image at different box sizes, and computes the fractal dimension as the slope between number of boxes and box sizes [10]. Two additional features were calculated: The range of the fractal dimension values across the total number of measurements taken, and the standard deviation of the different fractal dimension values across the various measurements taken.

All the above features were extracted for the image pairs in the recognition tasks from the dataset. These were then to be entered into the machine learning (random forest) classifier.

C. *Random Forest Learner and Classifier*

A random forest learner was implemented, with a 3-fold cross validation, to map the feature sets to the human-perceived complexity labels.

An important characteristic of the random forest method is that it yields feature importance information. The importance values were generated for each feature in Matlab as part of the *TreeBagger* function. Each value was calculated using the out of bag permuted predictor delta error, where a larger error corresponds to a more important feature for the mapping performance, in our case, to the complexity labels [11].

III. RESULTS

The subsequent results, obtained from following the above methodology, are presented in this section through an analysis on the data provided, and the importance levels of the features extracted from the images.

The response-times distributions change with subjects’ age, gender and computer skills are illustrated in Figures 9, 10, and 11, respectively. The graphs demonstrate that while the response time volumes vary within the factors, the distributions are similar across the age groups, between the genders, and across the computer skills. The response times could thus be assumed as an un-biased representation of the perceived complexity of the tasks performed by the subjects.

A. *Feature Importance*

The importance values of the features that provided the most insight are provided in Table III.

Table III implies that object-specific features have lower importance values compared to the whole-image features. The highest importance values of the whole-image features are those from the Gabor filter and fractal dimension calculations,

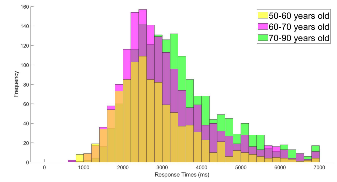


Figure 9. Frequency plot of the response times segmented by age

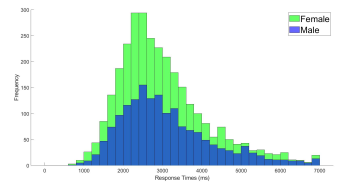


Figure 10. Frequency plot of the response times segmented by gender

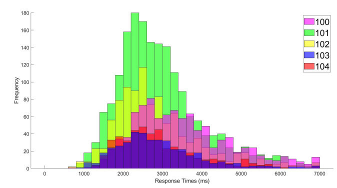


Figure 11. Frequency plot of the response times segmented by computer skill from 100 as computer illiterate to 104 as computer literate

as well as the white spacing feature and the symmetry. The strongest feature is the standard deviation of the Gabor filter. The smallest importance value for the whole image feature set, and one of the smallest in the entire feature set is the direct square comparison.

To illustrate the feature importance results of Table III, examples of recognition tasks from the dataset for which the complexity labels were correctly predicted by the algorithm are presented in Table IV. Three examples are given for each predefined complexity label: 1, 2 and 3.

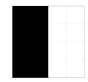
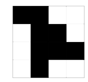
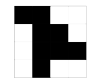
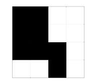
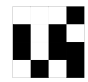
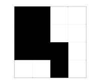
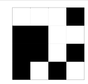
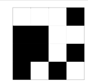
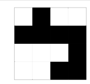
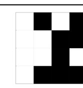
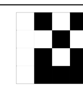
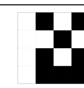
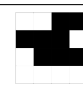
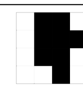
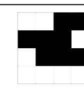
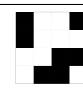
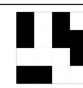
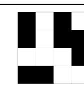
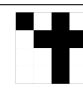
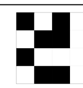
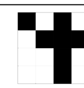
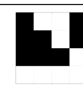
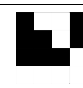
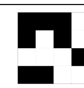
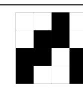
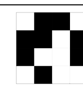
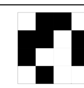
1) Object-Specific Feature Importance: The black adjacent path/s length/s yielded the highest importance value among these features. This can be explained by examples 2 and 3 of Table IV, where only one of the two different images has one long black adjacent path, with no other black objects. This makes it easily distinguishable from the other image in the task, and justifies the “easy,” label 1 mapping. The number of adjacent paths, which got low importance ranking, however, displays no specific trend in all tasks in Table IV across the three complexity labels. This visual trend can also be related to the relatively high importance value of the “Different black objects presence” feature: When one image in a pair has only one long black adjacent path and the other image has a variety of black features, the task was labelled as “easy” - label 1. The different white objects presence feature, however, does not display the same characteristic. Different white objects are present in several image pair examples of Table IV and across the various complexity labels, with no specific trend.

The importance values of the features for the objects’ presence and locations are all relatively small. The only object type whose location yielded a positive importance value was

TABLE III. PROMINENT FEATURE IMPORTANCE RESULTS

Whole-Image Features		Object-Specific Features	
Feature	**Importance Value**	**Feature**	**Importance Value**
Gabor filter std. dev	1.00	Black adjacent path length/s	0.73
Gabor filter sum	0.95	Different black objects present	0.73
Fractal dimension range	0.92	Similar black adjacent path location	0.61
White object spacing	0.86	Black similar objects within an image	0.57
Relaxed symmetry	0.86	White similar objects within an image	0.55
Fractal dimension	0.76	Number of black adjacent paths	0.41
Fractal dimension std. dev.	0.56	Number of white adjacent paths	0.39
Black object spacing	0.53	Similar white diagonal path location	0.08
Black similar objects in an image pair	0.42	Different white objects present	0.00
White similar objects in an image pair	0.26		
Direct squares comparison	0.23		

TABLE IV. EXAMPLES WITH CORRECT COMPLEXITY LABEL PREDICTIONS

#	Image 1	Image 2	Image 3	Complexity Label
1				1
2				1
3				1
4				2
5				2
6				2
7				3
8				3
9				3

the black adjacent path. In examples 5, 6 and 9 of Table IV, the presence of a similarly located black adjacent path in both images could have made them harder to distinguish between, thus increasing the perceived complexity in some way. Similarly, features related to similar shapes within an image yielded small importance values.

2) Whole-Image Feature Importance: The white object spacing feature had a much larger importance value compared to the black object spacing feature. In example 2 of Table IV, the average distance between the **white** objects of image 1, and image 3, is greater than that of image 2: There is only one long black adjacent path in the middle of image 1, whereas there are smaller black objects dispersed around image 2. This noticeable difference in the white spacing could have justified the ‘easy’, label 1, mapping.

The relaxed symmetry feature had relatively high importance value. In example 1 of Table IV, for instance, image 1 is symmetrical across the x axis, as well as the y axis, in inverted colours. This symmetry can explain the image as perceptually easy to discriminate in the recognition task, and could justify the “easy”, label 1, mapping for that task.

The feature of the direct squares comparison had no importance in the complexity predictions. This feature had no correlation to the human complexity labels with values randomly ranging from 4 to 12 for the tasks in Table IV.

The mathematical models - Gabor filters and fractal dimensions yielded the largest importance values. In the recognition tasks of Table IV, the values of all five features associated with these models: fractal dimension, range of the fractal dimension, standard deviation of the fractal dimension, sum of the Gabor filters and standard deviation of this sum, consistently decrease with increasing complexity.

IV. DISCUSSION

The feature importance analysis implies that whole-image features had greater significance than object-specific features. As per Witkin’s [3] field dependence concept, individuals may assess their visual fields as a whole, making only loose partitions of an image. This observed trend in our data may, however, be also related to the small size and simplicity of the images used, where individuals may not need to look for finer details.

The high importance of the number of black adjacent paths is in line with Attneave's [4] theories on visual redundancy, where long strings of adjoining pixels of the same colour could constitute for a level of visual redundancy, thus simplifying the visual-field's perceived complexity.

The small size of the images may also explain the result that the black spacing feature had a small importance value: The probability of having only one black object in these images is large - a fact that was assessed by the square counting feature. In this case, the average black object spacing feature is reduced to zero, which may produce a bias in the image pairs' black object spacing calculations. The higher importance of the average spacing of the white objects may be explained by the presence of more white objects in the image and is illustrated in the examples of Table IV, in the higher perceived complexity.

All white object-specific features demonstrated low importance values. The highest importance value related to white objects was the white object average distance feature, which was similar in strength to the black object average distance. This may be explained by associating the white objects distances with the black objects placements. These findings may imply that the white objects are perceived as background space.

The only whole-image feature that showed no importance in the complexity mapping was the direct squares comparison. This suggests that this typical computer process of scanning an image and comparing images is alien to human perception.

The significant importance of image symmetry in the complexity predictions, is in line with Attneave's [4] premise on image symmetry representing another form of visual redundancy, thus reducing the visual observation's perceived complexity. It indicates that when subjects observe an image that has symmetry, the mirroring makes it simpler to distinguish against other, non-symmetrical images.

Finally, the positive importance values of the mathematical Gabor filter and fractal dimension features imply their relation to visual perception. These importance values, and the trends showed that the smaller the **difference** in the values of these features for each of the two images in the pair, the greater the perceived complexity. This is also in line with the assumption in this study, that the smaller the difference between the images in the task, the harder it is to recognise the odd one. The feature importance trends found in the study thus suggest a potential of linking human visual perception to computational models of complexity.

While one might argue that the feature importance results have been obtained for images that are too small or simple, the motivation behind this chosen image set is that the lack of colour, and size of the images would allow for a more objective visual analysis by the subject. The disadvantage with presenting elaborate and natural visual stimuli, is that objects in the images, or the images themselves, are likely to visually trigger past memories or associations for a subject. These triggers could distract their focus off the required objective comparison of the images, thus skewing the human results that the algorithm is trying to mimic. It could, however, be worthwhile to explore designing stimuli that apply to real-world situations while preventing distractions in the future.

Additionally, many of these feature observations have touched on either Attneave [4] or Witkin's [3] theories at a high-level to demonstrate correlation to psychological research. This was only done at a high-level since the primary scope of this study was in trying to understand how computational or mathematical aspects of these images could be used in defining their visual complexity. However, in future work, it could be useful to conduct further research into psychological studies that have relevance to this topic. Given this, there is also potential for collaboration with psychologists to try link these results to human psychological perception mechanisms, and gain some insight to contribute to psychological research.

V. Conclusion

A method for the computation of visual tasks complexity combining information theory, machine vision and human perception measures was developed and assessed by its relevance to human perception. The computational complexity features could explain human visual complexity perception in the context of cognitive tests where this complexity perception is assumed to be represented by subjects' response times to visual tasks. The feature importance values corroborated psychological studies of human visual perception. These findings indicate a potential to link computer-extracted features to human perception in their definition of complexity of visual tasks.

References

[1] S. M. Anstis, "What does visual perception tell us about visual coding," Handbook of psychobiology, 1975, pp. 269–323.

[2] J. Wagemans, J. H. Elder, M. Kubovy, S. E. Palmer, M. A. Peterson, M. Singh, and R. von der Heydt, "A century of gestalt psychology in visual perception: I. perceptual grouping and figure–ground organization." Psychological bulletin, vol. 138, no. 6, 2012, p. 1172.

[3] H. A. Witkin and D. R. Goodenough, "Field dependence and interpersonal behavior." Psychological bulletin, vol. 84, no. 4, 1977, p. 661.

[4] F. Attneave and M. D. Arnoult, "The quantitative study of shape and pattern perception." Psychological bulletin, vol. 53, no. 6, 1956, p. 452.

[5] R. Mahendran, G. Jayashree, and K. Alagusundaram, "Application of computer vision technique on sorting and grading of fruits and vegetables," J. Food Process. Technol, vol. 10, 2012, pp. 2157–7110.

[6] V. Aharonson and A. D. Korczyn, "Human–computer interaction in the administration and analysis of neuropsychological tests," Computer methods and programs in biomedicine, vol. 73, no. 1, 2004, pp. 43–53.

[7] J.-K. Kamarainen, "Gabor features in image analysis," in 2012 3rd International Conference on Image Processing Theory, Tools and Applications (IPTA). IEEE, 2012, pp. 13–14.

[8] S. Marčelja, "Mathematical description of the response of simple cortical cells," Journal of the Optical Society of America, vol. 70, 12 1980, pp. 1297–300.

[9] Y. Fisher, E. Jacobs, and R. Boss, "Fractal image compression using iterated transforms," in Image and text compression. Springer, 1992, pp. 35–61.

[10] M. Bouda, J. S. Caplan, and J. E. Saiers, "Box-counting dimension revisited: presenting an efficient method of minimizing quantization error and an assessment of the self-similarity of structural root systems," Frontiers in plant science, vol. 7, 2016, p. 149.

[11] W. Man, Y. Ji, and Z. Zhang, "Image classification based on improved random forest algorithm," in 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), April 2018, pp. 346–350.