Automated Scoring of Bender Gestalt Test Using Image Analysis Techniques

Momina Moetesum*, Imran Siddiqi*, Uzma Masroor† and Chawki Djeddi ‡
* Center of Computer Vision and Pattern Recognition
Bahria University, Islamabad, Pakistan
Emails: reach.momina@gmail.com, imran.siddiqi@bahria.edu.pk
† Dept. of Professional Psychology, Bahria University, Islamabad, Pakistan
‡LAMIS Laboratory, Larbi Tebessi University, Tebessa, Algeria.

Abstract—**Drawing tests have been long used by practitioners and researchers for early detection of psychological and neurological impairments. These tests allow subjects to naturally express themselves as opposed to an interview or a written assessment. Bender Gestalt Test (BGT) is a well-known and established neurological test designed to detect signs of perceptual distortions. Subjects are shown a number of geometric patterns for reconstruction and assessments are made by observing properties like rotation, angulations, simplification and closure difficulty. The manual scoring of the test, however, is a time consuming and lengthy procedure especially when a large number of subjects is to be analyzed. This paper proposes the application of image analysis techniques to automatically score a subset of hand drawn images in the BGT test. A comparison of the scores reported by the automated system with those assigned by the psychologists not only reveals the effectiveness of the proposed system but also reflects the huge research potential this area possesses.**

I. Introduction

Neuropsychological testing represents a specialized area of assessment within clinical psychology. Psychologists use the contextual information provided by these tests for many purposes, for instance, the assessment of a child exhibiting academic problems or an adult with a degenerative brain illness. Medical procedures such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI), and Positron Emission Tomography (PET) scans can show the part of brain that seems abnormal, but neuropsychological tests show how a specific part of the brain is actually functioning.

Some neuropsychological testing procedures are lengthy and comprise of comprehensive batteries including a broad array of subtests like the Halstead-Reitan Neuropsychological Test Battery (HRB) [1] and the Luria-Nebraska Neuropsychological Battery (LNNB) [2]. Other neuropsychological tests are much briefer and are typically used as screens for neuropsychological impairment rather than full-fledged assessment tools. Examples include the Bender Visual-Motor Gestalt Test (BGT) [3] and the Rey-Osterrieth Complex Figure Test (ROCF) [4]. Nevertheless, a common feature amongst these tests is the use of figure or shape drawing. Poor performance on these tests is indicated by a variety of errors that subjects may make in copying the figures, including missing details, collision or overlapping, inability to accurately complete shapes like circles or squares, disproportionate size and angles or orientation. The patterns of scores on these subtests can go a long way towards pinpointing specific cognitive weaknesses.

Currently these tests are manually administered, scored and interpreted. This makes scoring and interpretation of large amounts of test data a time consuming and lengthy procedure [5]. Another significant factor that criticizes the use of such assessments is the biasness of a practitioner towards a subject belonging to a particular group [6]. The basic premise is that a screening device should not have an adverse impact on the overall assessment of the patient due to extrinsic factors.

Automation of a complete test or portions of a battery test can address most of the above mentioned problems. An automated system of scoring and interpretation of image based tests will enhance efficiency by reducing load of psychologists. Non-medical personnel can be trained to use this system and provide useful outcomes to the clinician. In this way, psychologists can focus more on the patient and treatment instead of scoring and report formulation. Such automated systems can also facilitate distant treatment where psychologists can recommend screening tests of a potential patient prior to conducting actual face-to-face session with them. The notion is applicable yet there are a number of challenges regarding the degree of accuracy of such a system [7]. The greatest challenge is the lack of precision in the drawing itself. A psychological patient is mostly unable to reconstruct the precise shape of an intended object or he/she may not be able to comprehend the image enough to reconstruct it. Therefore careful feature extraction is required to overcome some if not all of these problems. Nevertheless a fully or partially automated system can efficiently process large numbers of such drawings and the suspected cases can later be analyzed by human experts.

This paper presents the application of image analysis techniques for automatic scoring of a subset of drawings in the Bender-Gestalt test. Initial results of the study along with a comparison with human scoring are also presented. The paper is organized as follows. Section II reviews the previous work on similar problems while Section III describes the methodological implementations of the proposed system. Analysis of results obtained by the system and their comparison with manual scores are presented in Section IV while conclusions and future directions are highlighted in Section V.

II. Related Work

Early detection of neurological diseases through automated analysis of handwriting has been investigated by the document recognition community in a number of studies [8], [9], [10].

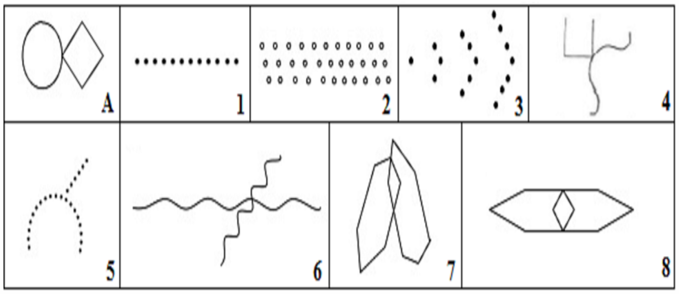


Fig. 1. Figures of Bender Gestalt Test

The automated analysis of hand-drawn images for detection of neuropsychological impairments, however, remains a relatively less explored area. In [11], authors have tried to automate the Rey-Osterrieth Complex Figure Test (ROCF) for evaluation of visual perception and long term visual memory. At the same time, automation of segments of popular drawing tests was also considered, in order to assert the potential scope of the given research problem. In [12], authors proposed a method of automatically segmenting a hand-drawn image into its components and used their location to establish the presence of constituent shape components. By determining the spatial relationship of the drawings, a series of standardized performance scores was produced to detect presence of visuo-spatial neglect. Further enhancement of their work is presented in [13] where constructional sequences from neuropsychological drawings are automatically assessed using Hidden Markov Models to obtain a diagnostic classification of patients with visuo-spatial neglect. Authors in [14] used hand-drawn images of simple three dimensional geometric shapes, such as a cube, to automatically detect early signs of Alzheimer's disease. More recent attempts in this field include the automation of a variation of the Clock Draw Test (CDT) [15] which is a screening test for people with cognitive impairments and dementia. Another technique for potential objective assessment of visuo-spatial ability, is presented in [16].

Most of the aforementioned studies involve one particular figure or similar segments of different figures in an effort to achieve automation. In this paper we propose automation of a subset of the widely used Bender Visual-Motor Gestalt Test (BGT) which comprises of a set of nine different figures as shown in Figure 1. In practical psychology, BGT is used to evaluate the visual-motor maturity and perceptual distortions associated with various neurological disorders [17], [18], [19], [20], [21]. The test comprises of showing a set of figures to the subjects which they are required to reproduce on a sheet of blank paper. The produced drawings are then analyzed to determine the presence or absence of certain properties like rotation, retrogression, angulation, simplification and closure difficulty, which assess the neurological state of the subject. A number of scoring systems for BGT have been proposed but the most popular and widely practiced scoring system is the Lack's scoring system, presented in [22]. Our proposed system will also evaluate the performance of hand drawn images on the basis of Lack's scoring system.

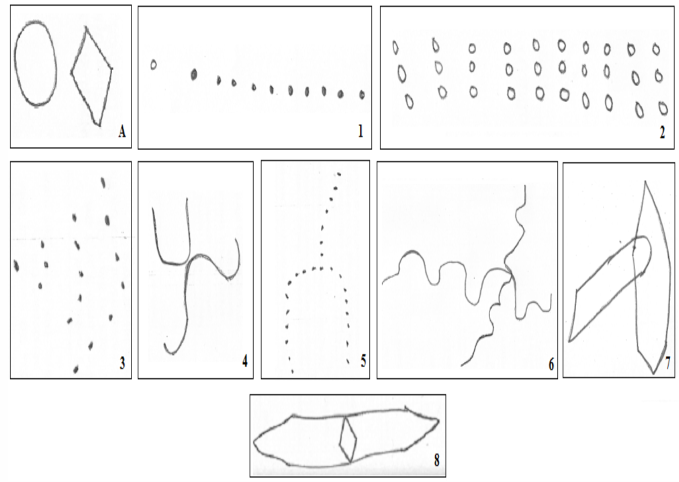


Fig. 2. Sample of figures drawn by a subject

III. METHODOLOGY

This section details the overall methodology of our proposed system. It highlights the data acquisition process, the scoring adaptations involved and the implementation of automated scoring.

A. Data Acquisition

A total of 152 drawings produced by 18 different subjects varying from 16 to 66 years of age are considered in our study. Samples were provided by the Department of Psychology at Bahria University, Islamabad. All the samples were acquired under expert supervision of trained psychologists. The hand drawn images on paper were then scanned to obtain the offline samples for processing. Figure 2 illustrates scans of drawings produced by one of the subjects involved in the test.

B. Scoring Properties and Automated Scoring

According to Lack's scoring system, each gestalt figure is analyzed individually against a set of scoring properties. Presence of these properties in a particular figure indicate presence of error. Table I shows the scoring scheme for each figure. Overall scoring is done based on the presence of error and not on the frequency of error i.e. if same property is observed in multiple figures only one mark is scored against it. Three or less than three errors indicate absence of any visuo-constructive deficits or brain impairments. Four errors are considered as a borderline score while five or more errors indicate presence of some brain impairments. Increasing number of errors is a strong indicator of presence of neurological disorders.

In some cases, during manual scoring, properties like time spent etc. are also taken under consideration which naturally cannot be handled in case of offline images. In addition, for computerized analysis of these scanned images, few adaptations were made after expert opinion of psychologists. The figures presented to the system are first binarized using global thresholding. Morphological closing is then applied to each figure to remove minor distortions, fill holes and get smooth object boundaries. Each figure is then separately analyzed and

TABLE I. SCORING PROPERTIES FOR GESTALT FIGURES

Scoring Properties	Gestalt Figures
Simplification	1, 2, 3 & 5
Overlapping Difficulty	6 & 7
Rotation	All figures
Perseveration	1 & 2
Closure Difficulty	A, 4 & 7
Cohesion	All figures
Angulation	2 & 3
Retrogression	All except 4 & 6
Motor In-coordination	All figures
Impotence	All figures
Fragmentation	All figures

scored while the final score is determined by combining the scores of individual figures. The scoring properties and their adaptations considered in our study along with the implementation details of their scoring are listed in the following.

1) Simplification: Simplification is marked for replacement of circles with dots or vice-versa and is applied to figures 1,2,3 and 5. To discriminate between dots and circles, we find the euler number (number of objects minus number of holes) of each connected component in a figure. Circles produce an euler number equal to zero while dots produce an euler number of 1. Since this parameter is sensitive to noise and distortions, the decision between circles and dots is carried out on the basis of a majority vote (of the individual components) and the figure is scored accordingly.

2) Overlapping Difficulty: Failure to overlap at the specified place or inability to overlap at all is termed as overlapping difficulty and is applicable to figure number 6 and 7. The complete inability to overlap can be determined by finding the number of components in each figure which should be 1 for overlapped figures. In case of overlapped figures, to find the overlap position, for figure 6, we find the skeleton of the image and find the branch points (having more than 2 neighboring pixels) and end points of the skeleton. In case of more than one branch points, the average of these points is considered as the intersection point. The intersection point is then joined with each of the end points (Figure 3) and the ratios of lengths of these lines are used to find abnormal overlap points. Similarly, for figure 7, conditional hole filling is carried out to fill the overlapped region. A series of set operations is then applied to segment the overlapped regions and the two shapes (Figure 4). The ratio of the area of overlapped region to the areas of each of the separate regions is finally computed and compared with predefined thresholds to identify unusually small or large overlap.

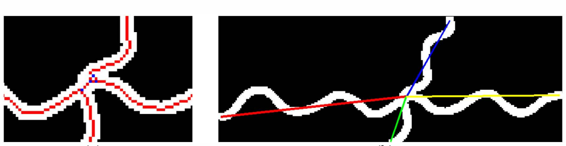


Fig. 3. Computation of intersection point (a): Skeleton of image and the interconnection points (b): Lines joining intersection point and the end points

3) Rotation: Rotation is marked for 80° to 180° orientation in a figure, no mark is assigned for lesser orientations. This property is the most critical for figure A where swapping of circle and diamond is a commonly observed phenomenon. We first apply hole filling algorithm to the image followed by erosion. This results in segmenting the circle and the diamond an example of which is illustrated in Figure 5. If the subject has drawn these objects in such a way that they do not touch each other, this step will have no impact on the image. Since hand-drawn shapes are, in general, crude, techniques like Hough transform to find circles or polygons may not work on these images. We, therefore, discriminate between circle and diamond by computing the compactness of each of the two objects in the image using the following circularity measure.

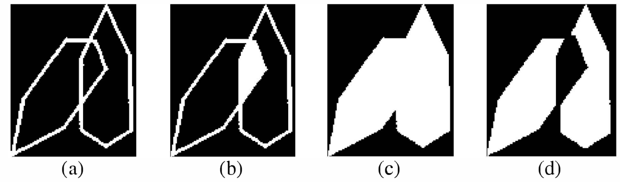


Fig. 4. (a): Original shape (b): Conditional hole filling (c): Completely filled shapes (d): Separated shapes without overlapped region

$$circularity = perimeter^2/(4 \times \pi \times area) \tag{1}$$

The object with higher compactness is considered a circle. The orientation of the line connecting the center of gravity of the circle to that of the diamond (Figure 5-c) is then used to score the rotation property.

For figures 1 and 2, we find the centroid of each connected component in the image and fit a least-square line to these points. The slope of this line is used to compute the orientation of the shape and the figure is scored accordingly. Since the figures are symmetric, the score is assigned for orientations $100° \leq \theta \leq 80°$.

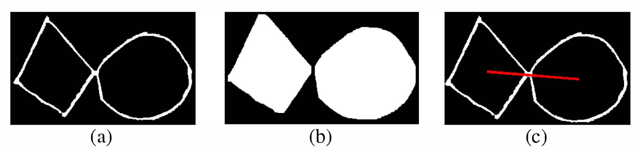
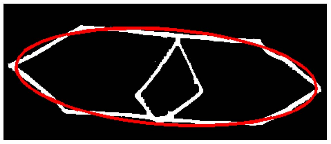


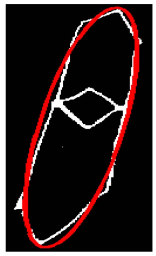
Fig. 5. (a): Original Image (b): Image after hole filling followed by opening (c): Line joining centers of gravity of two shapes

Orientation of figure 8 is computed by filling the shape (using region/hole filling algorithm) and finding the angle between the x-axis and the major axis of the ellipse that has the same second-moments as the shape. Examples illustrating fitting of an ellipse to compute the shape orientation can be seen in Figure 6. The scoring of rotation for figure 3 to figure 7 is presently under study and will be reported in a later communication.

4) Perseveration: Intra-design perseveration is marked if a figure is drawn beyond the limits specified by the stimulus. This property is checked for figures 1 and 2. In case of figure 1, more than 14 dots is considered an error. Similarly for figure 2, a maximum of 13 columns of circles must be present. Preservation is implemented by counting the total number of connected components in a figure and comparing it with the specific threshold of that figure. Isolated pixels and components whose area is below a predefined threshold are filtered prior to counting to eliminate noise and distortions.



(a)



(b)

Fig. 6. Ellipse fitting to figure 8 to compute orientation

5) Closure Difficulty: Difficulty in getting the joining parts of figures together is considered as closure difficulty and is checked for figures A, 4 and 7. From the view point of implementation, failure to join parts of figures lead to multiple connected components in the image. The number of components in the binarized and morphologically smoothed image of each figure is counted and the figure is scored accordingly.

6) Cohesion: An error in the cohesion of a figure is considered if there is an increase or decrease in the size of that figure by 1/3 of the dimensions used in the other drawings. The size of a figure is determined as the area of the bounding box encompassing the figure. The area of each figure is then compared with the median area of all figures to identify unusually small or large figures (difference of a scale factor of 3).

The aforementioned scoring properties represent a subset of the complete Bender-Gestalt scoring attributes. Other properties which have not been considered in our present work incclude angulation (variability of angulation of circles in figure 2), impotence (crossing off figures and redrawing attempts), motor incoordination (tremor-like lines or increased pressure at various points) and fragmentation (incomplete or broken up figures) and retrogression (substitution of a geometric shape by a different one). These properties will be considered and implemented in our further work on this problem.

IV. Results

This section presents the details of the experiments carried out to evaluate the proposed scoring methodology. The scores produced by the system for different figures and properties considered in our study are compared with the scores assigned through human inspection. A total of 152 figures produced by 18 different subjects were used for automatic scoring and the realized results are summarized in Table II ('-' refers to 'not applicable' while 'x' refers to 'not implemented'). It can be observed from these results that the system correctly scores a significant proportion of figures for almost all the properties. The only exceptional case is scoring of figure 6 for 'overlap' property where only 6 out of 12 drawings were correctly scored. This can be attributed to the fact that the property is scored based on finding the junction points on the skeletal image. The computation of these junction points is sensitive to image distortions and noise resulting in a false intersection point that eventually leads to a false score for the drawing. The errors in scoring of the 'simplification' were mainly due to drawings where the subjects used a mixture of dots and circles in a single drawing hence leading to a false score.

In general, the initial results are promising and can be further improved by introduction of more robust image analysis techniques to handle distortions and variations.

V. Conclusion

This paper presented our preliminary work on automation of a well-known neuropsychological screening test, the Bender-Gestalt test. We applied a set of image analysis techniques mostly based on morphological image processing, to automatically score different properties of this test. The initial results of automatic scoring are very promising and are likely to enhance further. The problem is quite challenging in the sense that the figures sketched by a subject are unconstrained. It should be noted that in our implementation, we have assumed that the drawings produced by subjects do not deviate from their respective prototypes to an extent where there are not recognizable. In situations where there is no correlation at all between a sketched drawing and the actual prototype, naturally, a human expert has to intervene.

Presently, we have implemented a subset of scoring properties and we intend to automate the complete set of properties in our further work on this subject. This will allow drawing meaningful conclusions about the subjects being tested. In addition, in our current system, different figures are separately evaluated. A strategy to automatically extract and identify different figures produced by a subject on a single page also needs to be devised. We also plan to work on online hand-drawn figures where the subject will be required to produce drawings on a digitizing tablet. This will allow access to additional information like time spent, pen pressure, speed and other dynamic parameters which will eventually lead to a better characterization of the subject.

Acknowledgment

This work was carried out as a pilot study for a project supported by the Franco-Pakistani Hubert Curien Partnership (PHC) - PERIDOT, implemented in Pakistan by the Higher Education Commission (HEC) and in France, by the Ministry of Foreign Affairs (MAE).

References

[1] R. M. Reitan, "Ward halstead's contributions to neuropsychology and the halstead-reitan neuropsychological test battery." *Journal of clinical psychology*, 1994.

[2] A. D. Purisch and R. J. Sbordone, "The luria-nebraska neuropsychological battery," in *Advances in clinical neuropsychology*. Springer, 1986, pp. 291–316.

[3] L. Bender, "A visual motor gestalt test and its clinical use." *Research Monographs, American Orthopsychiatric Association,*, vol. 3, no. xi + 176, 1938.

[4] M. Shin, S. Park, S. R. Park, S. Seol, and J. S. Kwon, "Clinical and empirical applications of the rey–osterrieth complex figure test," *Nature protocols*, vol. 1, no. 2, pp. 892–899, 2006.

[5] W. J. Camara, J. S. Nathan, and A. E. Puente, "Psychological test usage: Implications in professional psychology," *Professional Psychology: Research and Practice*, vol. 31, no. 2, p. 141, 2000.

TABLE II. RESULTS OF AUTOMATED SCORING

Figure Number	Total Figures	Number of Correctly Scored Figures					
		Simplification	Overlap	Rotation	Presrvation	Closure	Cohesion
A	18	-	-	15	-	15	17
1	18	14	-	18	16	-	15
2	18	16	-	18	17	-	15
3	18	15	-	x	-	-	14
4	16	-	-	x	-	15	14
5	18	16	-	x	-	-	14
6	12	-	6	x	-	-	11
7	18	-	14	x	-	16	16
8	16	-	-	15	-	-	14

[6] D. Burgess, M. V. Ryn, J. Dovidio, and S. Saha, "Reducing racial bias among health care providers: Lessons from social-cognitive psychology." *Journal of General Internal Medicine*, vol. 22, no. 6, p. 882887, 2007.

[7] L. Richard, "Continuing challenges for computer-based neuropsychological tests," *NeuroToxicology*, vol. 24, pp. 479 – 489, 2003, eighth International Symposium on Neurobehavioral Methods and Effects in Occupational and Environmental Health, Brescia, Italy, June 23-26, 2002.

[8] E. Kalcheva and G. Gluchev, "Segmentation and analysis of handwritten scripts from patients with neurological diseases," in *In Proc. of the 4th International Conference Conference on Computer Systems and Technologies: E-Learning*, ser. CompSysTech '03. New York, NY, USA: ACM, 2003, pp. 272–277.

[9] A. Schrter, R. Mergl, K. Brger, H. Hampel, H. Mller, and U. Hegerl, "Kinematic analysis of handwriting movements in patients with alzheimer's disease, mild cognitive impairment, depression and healthy subjects," *Dementia and Geriatric Cognitive Disorders*, vol. 15, no. 3, pp. 132–142, 2003.

[10] R. Plamondon, C. O'Reilly, C. Remi, and T. Duval, "The lognormal handwriter: Learning, performing and declining," *Frontiers in Psychology*, vol. 4, no. 945, 2013.

[11] R. Canham, S. Smith, and A. M. Tyrrell, "Automated scoring of a neuropsychological test: The rey osterrieth complex figure," in *Euromicro Conference, 2000. Proceedings of the 26th*, vol. 2. IEEE, 2000, pp. 406–413.

[12] R. M. Guest, M. C. Fairhurst, and J. M. Potter, "Automated extraction of image segments from clinically diagnostic hand-drawn geometric shapes," in *Euromicro Conference, 2000. Proceedings of the 26th*, vol. 2. IEEE, 2000, pp. 440–446.

[13] R. M. Guest, S. Chindaro, M. C. Fairhurst, and J. M. Potter, "Automatic classification of hand drawn geometric shapes using constructional sequence analysis," in *2013 12th International Conference on Document Analysis and Recognition*, vol. 2. IEEE Computer Society, 2003, pp. 990–990.

[14] A. Hazell and S. Smith, "Towards an objective assessment of alzheimers disease: The application of a novel evolutionary algorithm in the analysis of figure copying tasks," 2008.

[15] A. Guha, H. Kim, and E. Y.-L. Do, "Automated clock drawing test through machine learning and geometric analysis." in *DMS*, 2010, pp. 311–314.

[16] M. A. Lones and S. L. Smith, "Objective assessment of visuo-spatial ability using implicit context representation cartesian genetic programming," *Genetic and Evolutionary Computation: Medical Applications*, pp. 174–189, 2011.

[17] S. Sheikhi, "Clinical use of bender-gestalt test in brain lesions diagnosis and its comparison with magnetic resonance imaging (mri)," *Journal of Urmia Nursing And Midwifery Faculty*, vol. 5, no. 1, pp. 15–21, 2007.

[18] A. A. Santos and L. M. Jorge, "Bender test with dyslexics: comparison of two systems of punctuation," *Psico-USF*, vol. 12, no. 1, pp. 13–21, 2007.

[19] R. A. Allen and S. L. Decker, "Utility of the bender visual-motor gestalt testsecond edition in the assessment of attention-deficit/hyperactivity disorder 1," *Perceptual and motor skills*, vol. 107, no. 3, pp. 663–675, 2008.

[20] R. B. Ferreira, C. F. Feil, and M. L. T. Nunes, "Bender visual-motor gestalt test in the children's clinical assessment," *Psico-USF*, vol. 14, no. 2, pp. 185–192, 2009.

[21] L. Keppeke, I. Cintra, and T. H. Schoen, "Bender visual-motor gestalt test in adolescents: Relationship between visual-motor development and the tanner stages 1," *Perceptual & Motor Skills*, vol. 117, no. 1, pp. 257–275, 2013.

[22] P. Lacks, *Bender Gestalt Screening for Brain Dysfunction*, ser. Personality Processes Series. Wiley, 1984.