Automated Scoring of a Neuropsychological Test: The Rey Osterrieth Complex Figure.

R.O.Canham* S.L. Smith and A.M. Tyrrell
Department of Electronics, University of York,
Heslington, York, England.
sls@ohm.york.ac.uk

Abstract

The Rey Osterrieth Complex Figure (ROCF) is a widely used neuropsychological test for visual perception and long term visual memory. Many scoring systems are used to quantify the accuracy of the drawings; these are currently implemented by hand in a subjective manner. This paper gives details of the current progress of a novel technique to locate the scoring sections of the most common of these system (the Osterrieth Scoring System), with the ultimate goal to automating the scoring system. High levels of distortion are possible making this an extremely difficult task; however, location and perceptual grading of the basic geometric features (triangles, rectangles and diamonds) have been most successful. All but one section in the test data was located (99.3% success) and 78% of the perceptual grades calculated were within 5% of grades generated by independent raters. Unary spatial metrics have been implemented to reduce the possible section candidates by an average of 75% without the loss of a single section.

1. Introduction

Neuropsychologists (concerned with the behavioural expression of brain dysfunction [19]) make use of many tests when assessing neurological dysfunction in a subject. Much information can be obtained by the use of advanced scanning techniques (such as CAT and MRI scans); however, there are still cases where a simple paper and pencil test can give additional, valuable information. In many situations the size of a lesion (a localised abnormal tissue change) does not accurately reflect the degree of dysfunction. Neuropsychological tests can provide valuable data concerning the progress of patients through treatment and provide a key tool for research into the organisation of brain activity and its translation into behaviour, brain disorders and behavioural disabilities. One such test is the Rey Osterrieth Figure (ROCF), which was devised by Rey [24] and standardised by Osterrieth [24] to test visual perception and long term visual memory function. It is used as a neurological evaluation tool for both children and adults for a diverse number of conditions from child developmental problems to dementia, trauma and infectious processes. The test requires the subject to copy, and later reproduce from memory the diagram shown in figure 1.

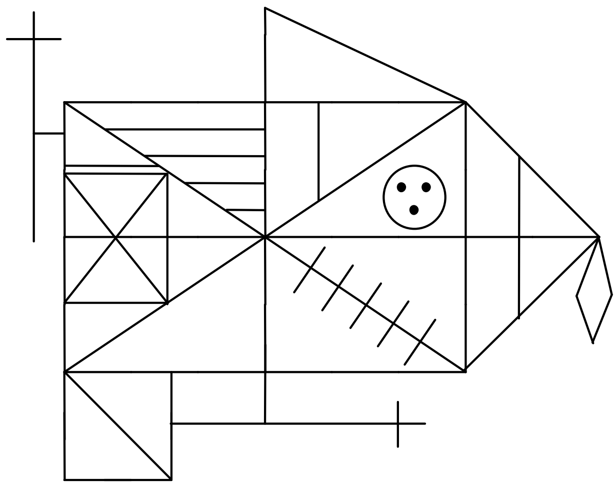

Figure 1. The Rey-Osterrieth Complex Figure. Typically 20cm in length.

The ROCF is widely used, both in a clinical and research environment and numerous studies have been performed upon it (see [19, 29] for extensive lists). The order and accuracy in which the figure is copied and drawn from recall provides useful information concerning the location and extent of any damage. To derive a more quantitative value for the accuracy of a subject's drawing various scoring systems

*Supported by the Engineering and Physical Sciences Research Council

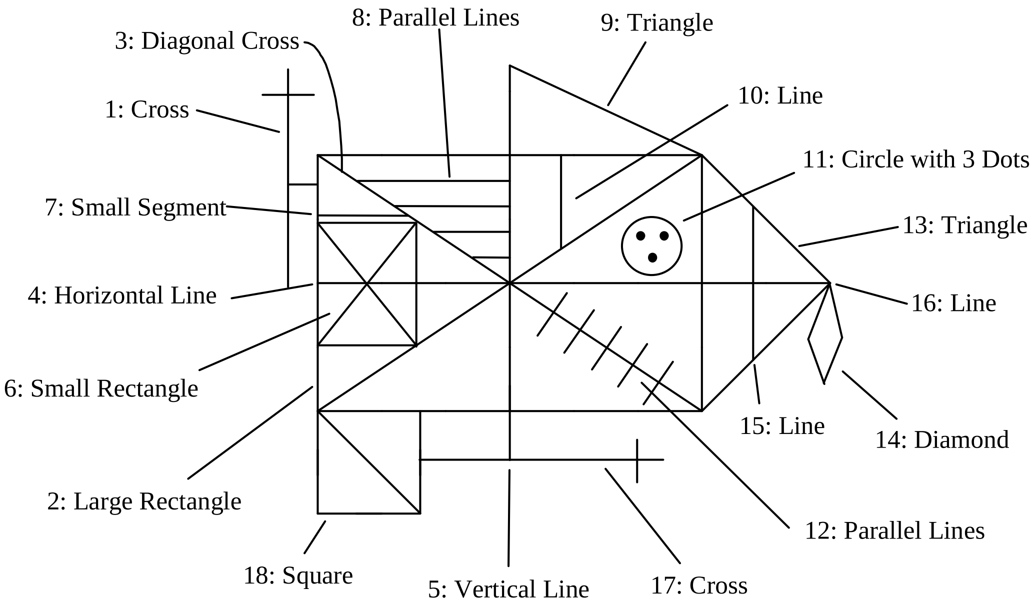

Figure 2. The Osterrieth Scoring System

Unit Correct	Placed Properly	2
	Placed Poorly	1
Unit Distorted, incomplete but recognisable	Placed Properly	1
	Placed Poorly	1/2
Absent or Unrecognisable		0

Table 1. Osterrieth marking allocation

have been employed. The most widely used is the Osterrieth system [19, 24]; the figure is split in to eighteen identifiable areas (see figure 2), each of which is considered separately and marked on the accuracy of its position and the distortion exhibited, using the scale shown in table 1.

Limitations of this scoring system include the lack of organisational information (such as whether the drawing was produced in a piecemeal or a logical fashion) and failure to differentiate the diagnostic importance of different sections. Consequently a number of scoring systems have been developed, including those by Waber and Holmes [31, 32], Bennett-Levy [2], Hamby [14], Fastenau [10] and The Boston Qualitative Scoring System by Stern et al [28].

These scoring systems are currently performed by hand in what tends to be a subjective manner, which is open to interpretation. The Osterrieth system was accurately defined by Taylor (reproduced in [27]); however this is not universally adhered to. The individual scoring sections tend to have very poor inter-rater reliability [29] and the system has been criticised for its lack of thorough testing [10]. It has also been noted that to aid marking, the criteria have been set artificially strict or lenient [2]. Various aspects of the Waber and Holmes' system have also produced low inter-rater reliability. The Boston Qualitative Scoring System makes use of guides and templates to produce a very comprehensive score; however, one drawing takes between five and 15 minutes to mark.

It is proposed to produce an automated implementation of the Osterrieth scoring system. Such an automation would not only provide an objective and consistent result but would also alleviate a highly skilled clinician from a tedious and time consuming task.

Recording a subjects' drawing using a digitising tablet would provide an unobtrusive method of recording the constructional sequence that is required in some scoring systems (no satisfactory process is currently available). However, the tablet also records dynamic data which has been shown to contain valuable information on simpler neuropsychological copying test [9], opening up an interesting avenue of research.

The first step in this automation is the location of the relevant scoring sections within an off-line, scanned image.

This paper gives an overview of the current progress of this work with a detailed description of the ROCF to place the work in context. Full technical details can be found in [6, 5] by the authors.

Section 2 of the paper gives a review of previous work and provides an overview of the difficulties of the problem. The proposed technique is given in section 3 and the results are shown in section 4. Finally the paper's conclusions are given in section 5.

2. Overview of Problems and Previous Work

Automating the location of the ROCF scoring sections is a very difficult problem. The ROCF is, by definition, a complex figure and the reproductions by patients typically have very high levels of distortion, many of which have clinical implications. The figure can be drawn in a piecemeal fashion, with sections misplaced, repeated or missing altogether. Sections can have large gaps in the sides or corners, be constructed using multiple strokes, be squashed or twisted with curved or stepped sides.

Hand drawn line figures and sketches are generated in a number of applications and a great deal of work has been performed to provide robust techniques to interpret them. Both on and off-line applications have been considered, including computer graphical user interfaces [7, 26] and conversions for CAD input or for tidying plans or schematics [3, 22, 4, 23, 13, 1]. These images contain an inherent degree of distortion and inaccuracy that the techniques must be able to accommodate, however, the distortion produced by the ROCF is beyond the capabilities of these systems. Neural networks [20, 30] and other adaptive/learning techniques [16, 11, 12, 17] have been applied to hand drawn figures and similar applications but with much simpler data sets with, by comparison to the ROCF, minimal distortion. With such a vast degree of variation possible in the ROCF it would be difficult to produce the large training set required for such systems and no suitable technique has been identified.

Although much recent effort in computer vision techniques has, understandably, been aimed at adaptive and automated systems, there is a large body of work that reduces an image into a complex line image before applying a suitable knowledge based system (see [8] for a comprehensive survey). Many make use of Gestalt psychology to identify features; a system that identifies perceptually significant grouping properties in human vision based upon features such as co-termination, continuation along a straight or smoothly changing path, symmetry and closure [15]. A survey of perceptual organisation in computer vision can be found in [25]. There is a significant difference between these computer vision systems and the application detailed here; the computer vision system is inherently probabilistic in nature, where a fundamentally correct image is distorted by noise, optical imperfections or problems associated with the feature extraction, while the ROCF data is inherently fuzzy in nature.

3. Approach Adopted and Implementation to Date

The approach taken to locate the scoring sections is to first identify all the suitable basic geometric shapes within the figure. In order to facilitate development, only scoring sections based on triangles, rectangles, diamonds and simple lines are to be considered at this time. However, this will still demonstrate the suitability of the techniques employed. With such large levels of distortion it is difficult to crisply categorise a shape as being present or not and so a grade of perceptual distortion is calculated which not only provides a scale to which a cut off point can be assigned but can also be used as a metric in further processing.

3.1. Basic Geometric Feature Location and Rating

The geometric features are located by the search of an Attributed Relational Graph (ARG) that is generated to represent the vectorised binary scanned image. ARGs have been used in similar applications such as [22, 21]. The vectorisation is a simple process of thinning the image and then applying a line following algorithm. Line segments that are collinear are then joined using a novel collinear metric based upon fuzzy metrics of the lines closeness and difference in angle [6].

The ARG is constructed to represent the connectivity of the collinear lines; each node of the ARG represent a collinear line and the joining arcs represent a connected line. However, due to the distortion the connectivity is often broken and so a 'closeness' metric is used. The ARG is searched independently for each shape. Each node is used as a starting point and all the connected nodes are tested. If the test succeeds that line becomes the current line and the process is repeated, while failure results in chronological backtracking to perform an exhaustive depth first search.

The test is in the form of a set of production rules that describes a corner, a line continuation and a corner truncation. The distortion can be so high that the lines constructing a straight side can have such a large deviation that under other circumstances it could be considered a corner and so the joining of line segments to form collinear lines can only be practised on strongly collinear lines. To accommodate continuous sides with greater distortions it is necessary to include a definition of a straight line continuation in the ARG search.

A common problem with the large rectangular scoring section (section number 2) is the truncation of a corner as shown in figure 3. Another rule set has been created to account for a truncated corner.

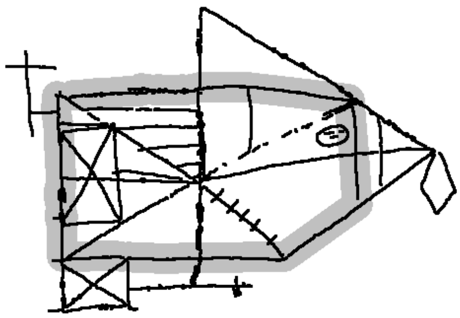

Figure 3. Example of a truncated rectangle

If the original node is encountered and the correct number of corners have been found then the shape data is stored.

Each candidate shape found is then ‘rated’ to describe its level of perceptual distortion. This is naturally described by fuzzy sets and linguistic variables [33] that allow an intuitive description of a shape to be produced based upon Gestalt principles of perceptually significant features. This calculates a membership to rate how “good” the shape is.

The metric used is the collinearity of each side, based upon the difference of angle and closeness of line termination. However, this gives a ‘local’ collinearity metric and so a global metric of the maximum perpendicular deviation from the straight path is also included. Corner properties of the deviation from 90° (for rectangles) and closeness of line termination are used together with a symmetry metric for diamonds. Each metric is combined using an appropriately weighted Yager intersection function [18]. Many of the metrics make use of relative sizes and distances (hence the rating process could not be integrated in to the location process since the size of the feature is required). However, it was found that relative measure had a tendency to score very large or very small structures either too leniently or harshly; there is a more complex relationship due to the figure’s overall size, and so an absolute element was also included. The parameters used to control the fuzzy metrics are currently set by hand; however, they are in a format that can be automated using a suitable genetic algorithm.

Due to the multi-stroke nature of the drawn figures it is possible for a number of very similar features to be found and categorised as seperate shapes. A fuzzy closeness metric is used to group similar features into a single element. For full technical details see [6].

3.2. Identification of Scoring Sections

The collection of geometric shapes must now be examined to identify the correct features. This is a considerable problem, since the variation of the figure is extreme. There is no guaranteed datum within the figure and most metrics must be relative to an uncertain base. The computational expense of calculating these relative metrics is great and so a first pass is performed to remove the most unsuitable features using unary metrics where only absolute features are considered. The goal of this process is to remove as many unsuitable features as possible without the loss of any scoring sections.

To locate the sections a number of basic spatial relations are considered, grouped in to the approximate categories of position, orientation, size and basic features. Once again fuzzy logic is employed to accommodate the distortion and variations possible in a natural and intuitive manner.

Basic Shape Features. The most fundamental basic shape feature is the shape type itself, whether it is a triangle, rectangle or diamond. Rectangles have an aspect ratio while triangles can be a right angled triangle and have symmetry in a given plane.

Size. The size of the feature is taken as the ratio of the feature’s area and the total figure area to give an indication of the size in both planes. However, the square root is taken to give a more linear attribute.

Position. The position of a feature is considered independently in the x and y plane since it is possible for it to be correctly positioned in one but not the other plane. All corner points are considered and the worse case used.

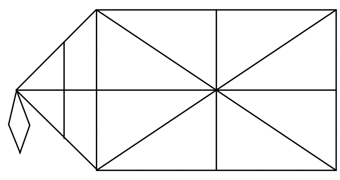

Figure 4. Example of an incorrectly placed triangle with symmetry

Orientation. The basic orientation of a feature is a function of its side angles compared to the orientation angle being considered. Diamond features used the angle of the axes. Triangles also have directional orientations. A triangle with a vertical side can be facing to the left or right. Using triangular section 13 in figure 2 as an example, it is clear to see that it should have a right facing orientation. Triangles with a horizontal side can also have an up and down facing orientation and a suitable right angle triangle

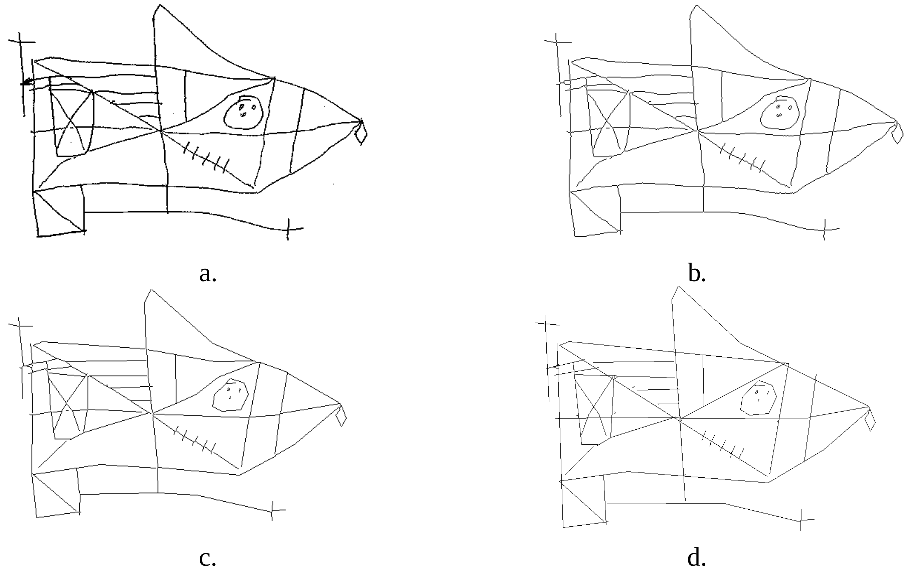

Figure 5. Example of pre-processing steps. a. original b. thinned image c. vector representation (consists of 241 lines) d. collinear lines (consist of 119 lines)

can have both a left/right and up/down orientation (see section 9 in figure 2).

The location of the features can be missed-placed but must still be identified. Symmetry is identified by Gestalt psychology [15] as an important factor. Hence a missplaced feature with symmetrical properties to its correct location must be considered as more significant that one without. Thus when considering the direction of a triangle, the direction from the centre line is also include, again in the appropriate plane. If the triangular section 13 is used as an example and placed in a symmetrical position as shown in figure 4 then it will face away from the centre line.

The metrics calculated for each section are aggregated to form a single measure using a generalised mean [18]. See [5] for further technical details.

4. Results

The techniques were tested using a random sample of 31 drawings of the ROCF produced by children attending the Institute of Child Health, London, who displayed a typical spread of illnesses seen by the neuropsychological unit. Of these, 16 drawings were produced by copying the figure and 15 from recall. Only the rectangular, triangular and diamond scoring sections were considered, which constitute scoring sections 2, 6, 9, 13, 14 and 18 (see figure 2).

Of the 31 drawings many had sections missing or so highly distorted that it is not possible to locate them at this stage in the processing and are only identifiable by a human observer with use of contextual information. Hence a total of 140 scoring sections were considered; their composition is shown in table 2.

Section number under consideration	2	6	9	13	14	18
Number of sections present	28	20	23	26	22	21

Table 2. Composition of scoring sections present

The pre-processing stages of thinning, vectorisation and grouping of collinear line segments performed well, reducing the mean number of collinear lines to 126 from 286 original lines. Some examples are given in figure 5.

The location process also performed very well, locating all sections except one (99.3% success). The nature of the distortion of the missed feature is such that it is very difficult to locate and hence it is best located in a later processing stage with the aid of contextual information. A selection of example scoring sections found are given in figure 6.

The perceptual grades calculated by the process were compared to scores generated by six independent raters. To simplify this grading process the raters where asked to grade each shape into a class that mapped onto a band of values within the automated scale. If the calculated grade fell outside the band then the distance to the edge of the band was expressed as a percentage error. The subjective results were quite diverse and so the modal average was taken to remove the extreme scores.

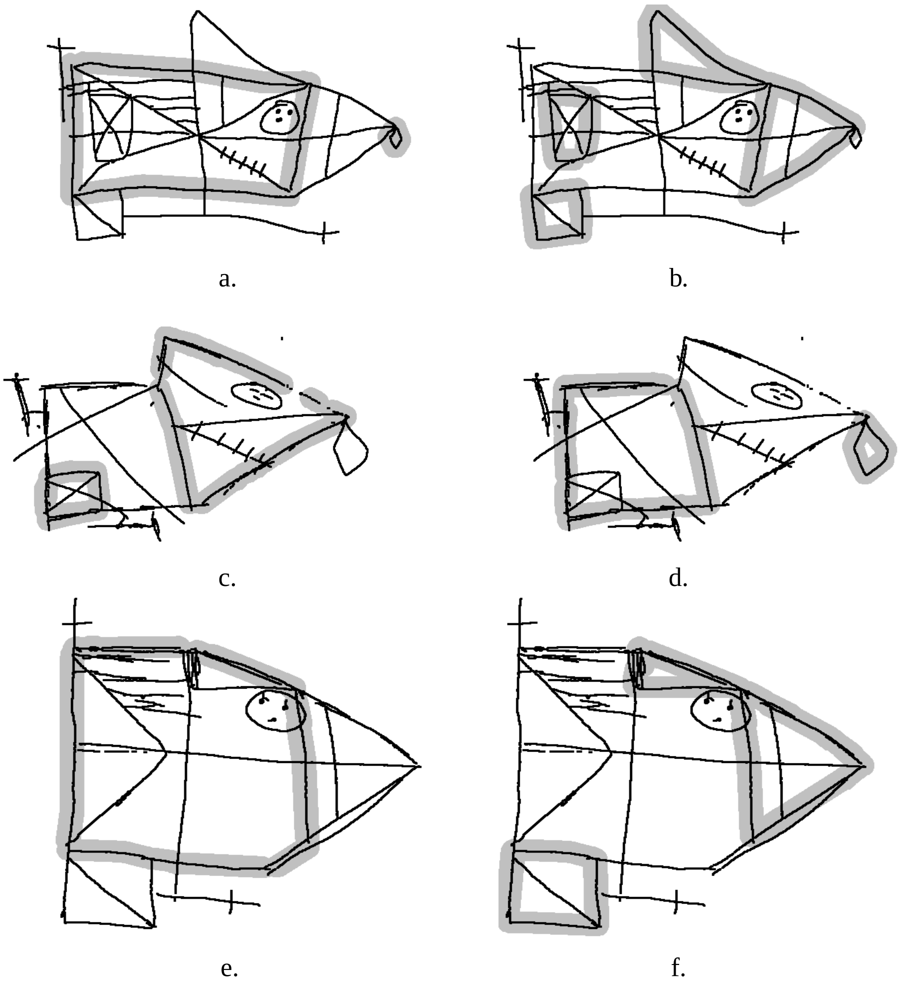

Figure 6. Example of scoring sections located. Rectangle 2 and diamond 14 are highlighted in (a), (c) and (e). Rectangle 6, triangle 9, triangle 13 and square 18 are highlighted (if present) in (b), (d) and (f).

The rating process did prove to generate very good approximations for the degree of distortion. When compared to the grades generated by the independent raters, some 78% of the calculated results had an error of 5% or less and all but 2 features (98.6% of the data) had an error of 10% or less. Table 3 shows the percentage of shapes with an error of 5% and 10%, or less, broken down into the individual scoring sections.

The average number of shapes found with a perceptual grade higher than the working threshold was 53.5 rectangles, 103.2 triangles and 48.8 diamonds per figure. The unary metrics were used to discard features that were unsuitable for each scoring section and hence discarded an average of 75% of these features without the loss of a single scoring section. The breakdown for the individual scoring sections is given in table 4.

It is noticeable that the rectangular scoring section 6 performed less well compared to the other sections. This is understandable when the data is examined; it is in a area of high line density with a high degree of variability in size and position possible for that feature. Hence it is not possible to discard too many features without danger of discarding a scoring section.

Section	2	6	9	13	14	18
Error $\leq$ 5%	75	75	77	81	77	86
Error $\leq$ 10%	100	95	100	100	95	100

Table 3. Percentage of features with calculated grades within given errors of independent raters' grades

Section number	2	6	9	13	14	18
Percentage reduction	68.0	50.4	71.6	81.8	89.0	88.9

Table 4. Percentage of features discarded, using unary metrics, for each scoring section

5. Conclusions

The Rey Osterrieth Complex Figure (ROCF) is a "pen and paper" neuropsychological test used to evaluate neurological dysfunction in visual perception and long term visual memory. A subject is asked to copy the complex figure and then reproduce it from memory. It is widely used in research and clinical environments. The Osterrieth scoring system is the most popular system of many scoring systems available that produce a quantitative score for the accuracy of the drawing. Currently the scoring is undertaken manually in a subjective manner and has been criticised for its unreliability in a number of publications. Automating the scoring process, as described in the paper, produces an objective result and removes a time consuming and tedious task from a skilled clinician (some schemes can take up to 15 minutes per figure).

The first stage of this automation is the ability to identify the scoring sections within the possibly highly distorted figure. A novel process, that employs fuzzy metrics based upon Gestalt psychology, has been described that locates and grades the basic geometric shapes on a scale of perceptual distortion. This process functioned extremely well, locating all but one feature from a random set of test drawings (99.3% success). The grading process also performed well when compared to subjective grades produced by 6 independent raters; 75% of the features were within 5% of the subjective grades and 98.6% within 10%. The process to identify the relevant scoring section from within all the geometric shapes found requires a computationally expensive process using binary metrics. A set of unary metrics have been implemented to remove unsuitable features and hence speed the binary metric calculations. This process removed an average of 75% of the features without removing a single scoring section.

. Acknowledgements

The authors are extremely grateful to Dr Elizabeth Issacs and her colleagues at the Institute of Child Health, London, for access to example patient responses to the ROCF, and their contribution to this work.

References

[1] A. Apte, V. Vo, and T. Kimura. Recognizing multistroke geometric shapes: An experimental evaluation. *Proc. 6th ACM Symposium on User Interface and Software Technologies*, 3:121 – 8, 1993.

[2] J. Bennett-Levy. Determinants of performance on the rey osterrieth complex figure test: An analysis and a new technique for a single-case assessment. *British Journal of Clinical Psychology*, 23:109 – 119, 1984.

[3] Y. Bin and Y. Baozong. A consistent attributed graph-based hand-drawn circuit diagram reading system. *Chinese Journal of Electronics*, 4(1):1–11, 1995.

[4] H. Bunke. Attributed programmed graph grammars and their applications to schematic diagram interpretation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 4(6):574–582, 1982.

[5] R. Canham, S. Smith, and A. Tyrrell. Fuzzy spatial metrics for the identification of features from within a highly distorted, complex line drawing. *BMVC 2000*. In Press.

[6] R. Canham, S. Smith, and A. Tyrrell. Recognition of severely distorted geometric shapes from within a complex figure. *Pattern Analysis and Applications*. In Press.

[7] C. Chen and S. Xie. Fuzzy freehand drawing system. *IEEE International Conference on Fuzzy Systems*, II(367):1012 – 1017, 1994.

[8] D. Crevier and E. Lepag. Knowledge-based image understanding systems: A survey. *Computer Vision and Image Understanding*, 67(2):161 – 185, 1997.

[9] M. Fairhurst and S. Smith. Application of image analysis to neurological screening through figure-copying tasks. *int J Biomed Compt*, 28:269 – 287, 1991.

[10] P. Fastenau, J. Bennett, and N. Denburg. Application of psychometric standards to scoring system evaluation: Is 'new' necessarily 'improved'? *Journal of Clinical and Experimental Neuropsychology*, 18(3):462 – 472, 1996.

[11] A. Fred, J. Marques, and P. Jorge. Hidden markov models vs syntactical modeling in object recognition. *Proceedings of the International Conference on Image Processing*, 1:892 – 6, 1997.

[12] R. Greenspan, R. Goodman, and C. Anderson. Learning texture discrimination rules in a multiresolution system. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(9):894 – 901, 1994.

[13] M. Gross. Recognizing and interpreting diagrams in design. *Proceeding of the Workshop on Advanced Visual Interfaces*, pages 88 – 94, 1994.

[14] S. Hamby, J. Wilkins, and N. Barry. Organisational quality on the Rey-Osterrieth and Taylor Complex Figure Tests: a new scoring system. *Psychological Assessment*, 5(1):27 – 33, 1993.

[15] E. Hilgard, R. Atkinson, and R. Atkinson. *Introduction to Psychology*. Harcourt Brace Jovanovich Inc, 6 edition, 1975.

[16] M. Katagiri and M. Nagura. Recognition of line shapes using neural networks. *IEICE - Transactions on Information and Systems*, 7:754 – 60, 1994.

[17] J. Keller and R. Kirshnapuram. Fuzzy decision models in computer vision. In R. Yager and Z. Zadeh, editors, *Fuzzy sets, Neural Networks and Soft Computing*, pages 213 – 232. Van Nostrand, 1994.

[18] G. Klir and T. Folger. *Fuzzy Sets, Uncertainty, and Information*. Prentice-Hall Internatinoal, 1988.

[19] M. Lezak. *Neuropsychological Assessment*. Oxford University Press, New York, 2 edition, 1983. pp395-402.

[20] Y. Liu. Encoding more difficult shapes using a neuron like model. *Neural Network World*, 7(6):729 – 38, 1997.

[21] J. Llados, K. Lopez, and E. Marti. A system to understand hand-drawn floor plans using subgraph isomorphism and Hough Transforms. *Machine Vision and Applications*, 10:150 – 158, 1997.

[22] B. Messmer and H. Bunke. Automated learning and recognition of graphical symbols in engineering drawings. *Lecture Notes in Computer Science: Graphics, Recognition, Methods and Applications. First International Workshop. Selected Papers*, pages 123–134, 1996.

[23] B. Pasternak. Processing imprecise and structural distorted line drawings by an adaptable drawing interpretation kernel. *International Association for Pattern Recognition systems*, pages 318 – 37, 1994.

[24] A. Rey and P. Osterrieth. Translations of excerpts from Rey's 'Psychological Examination of Traumatic Encephalopathy' and Osterrieth's 'The Complex Figure Test'. *The Clinical Neuropsychologist*, 7:2–21, 1993.

[25] S. Sarker and K. Boyer. Perceptual organization in computer vision: A review and a proposal for a classification structure. *IEEE Transactions on Systems, Man and Cybernetics*, 23(2):382 – 99, 1993.

[26] E. Saund and T. Moran. A perceptually-supported sketch editor. *UIST '94. The Seventh Annual Symposium on User Interface Software and Technology. Proceedings*, pages 175–84, 1994.

[27] O. Spreen and E. Strous. *A Compendium of Neurological Tests. Administration Norms and Commentary*. Oxford University Press, 1991.

[28] R. Stern, L. Singer, N. Duke, C. Morey, E. Daughtrey, and E. Kaplan. The Boston Qualitative Scoring System for the Rey Osterrieth Complex Figure: Description and inter-rater reliability. *The clinical Neuropsychologist*, 8(3):309 – 322, 1994.

[29] L. Tupler, K. Welsh, Y. Asare-Aboagye, and D. Dawson. Reliability of the Rey-Osterrieth Complex Figure in use with memory - impaired patients. *Journal of Clinical and Experimental Neuropsychology*, 17(4):566 – 579, 1995.

[30] F. Ulgen, A. Flavell, and N. Akamatsu. On-line shape recognition with incremental training using binary weights algorithm. *Applied Intelligence*, 6(3):225 – 240, 1996.

[31] D. Waber and J. Holmes. Assessing children's copy production of the Rey-Osterrieth Complex Figure. *Journal of Clinical and Experimental Neuropsychology*, 7(3):264 – 280, 1985.

[32] D. Waber and J. Holmes. Assessing children's memory production of the Rey-Osterrieth Complex Figure. *Journal of Clinical and Experimental Neuropsychology*, 8(5):563 – 580, 1986.

[33] L. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338 – 353, 1965.