Art Therapy: Journal of the American Art Therapy Association, 26(2) pp. 73-79 © AATA, Inc. 2009

Brief Report

A Computer System to Rate the Variety of Color in Drawings

Seong-in Kim and **Ibrahim A. Hameed,** Seoul, Korea

Abstract

For mental health professionals, art assessment is a useful tool for patient evaluation and diagnosis. Consideration of various color-related elements is important in art assessment. This correlational study introduces the concept of variety of color *as a new color-related element of an artwork. This term represents a comprehensive use of color, which is a trait that is usually subjectively assessed by a rater's personal knowledge, experience, feeling, and intuition. A sample of children's crayon drawings (*N *= 52) were evaluated both by human raters and also by a computer system that rated the variety of color present in each artwork by automatically detecting the number of colors used and the length of edges between colors. Comparing the human ratings with the computer ratings showed a high correlation of the results, leading to a conclusion that the computer system can be a useful aid for art assessment by human raters.*

Introduction

Art assessment is widely accepted as a valuable clinical technique for mental health professionals (Oster & Gould, 2004). Many art assessment tools have been developed for rating the various elements in a drawing so as to provide helpful information regarding a patient's diagnosis and emotional state. Historically, the use of color in a drawing has been considered an important a factor in art assessment, along with other elements such as theme, line, and form (Ghaffurian, 1995). Rorschach (1951) regarded color as a means for revealing a person's emotion. Some art therapists have reported that child victims of severe sexual abuse (Malchiodi, 1990) and patients suffering from depression (Gantt & Tabone, 1998; Wadeson, 1980) tend to use only one or two colors in their drawings. In one study, patients diagnosed with substance abuse used less color than patients in a comparison group (Francis, Kaiser, & Deaver, 2003). Children who have experienced natural disasters tend to use a limited number of colors (not more than two or three) and a palette mostly consisting of black, white, and sometimes red (Gregorian, Azarian, DeMaria, & McDonald, 1996). Survivors of trauma often express their psychological pain, anxiety, fear, sorrow, loneliness, and hopelessness by selecting particular colors, as do other patient populations.

A drawing contains various color-related elements that are believed to reveal the client's emotional status at the time of its creation. These elements include the number of colors used, the hues chosen, the prominence of color, the emotional tone, and the degree of color mix. Most art assessments include various color-related elements in their rating systems (Kim, Bae, & Lee, 2007). For example, the rating system of the Diagnostic Drawing Series by Cohen (1986/1994) incorporates color type, blending, and idiosyncratic use of color; the Formal Elements Art Therapy Scale rating system (Gantt & Tabone, 1998) includes prominence of color and color fit. We propose *variety of color* in drawings as a new color-related element in art assessment.

Drawings can be used to determine the progress of patient treatment (Betts, 2006). In art therapy, for instance, an assessment can be administered at the onset of treatment, during its middle phase, and again upon termination. However, the limitation common to all art assessments and the research underlying them involves the subjectivity of rating art. Raters often proceed on the basis of subjective and rather uncertain knowledge, relying on professional observation and judgment. Although raters usually are provided with concrete descriptors for their ratings, it is still conceivable that they may rate aspects of drawings differently simply because they like a certain drawing better than others (White, Wallace, & Huffman, 2004). All ratings of elements are more or less subjective and the results may differ depending on the raters.

Certain elements, such as variety of color, elaboration, and emotional tone, may be particularly difficult to rate objectively. Thus, a computer system programmed to serve as an objective rating tool would be of great value to human raters. For our study, we selected the variety of color as a

---

Editor's note: Seong-in Kim, PhD, is Professor of Industrial Engineering at Korea University, with a research interest in developing new expert systems for art therapy. Ibrahim A. Hameed is a PhD candidate at Korea University. This report was presented at the 2007 Annual Conference of the American Art Therapy Association, Albuquerque, NM. The research was supported by the Second Brain Korea 21 Project. Correspondence concerning this report may be addressed to tennis@korea.ac.kr

focal element for such an application. We define variety of color as the patient's comprehensive use of color when creating an artwork. This trait is usually subjectively assessed by a rater's personal knowledge, experience, feeling and intuition.

We propose an interdisciplinary study of art assessments and computer technologies. By applying technologies such as blurring (Gonzalez & Woods, 2002) and clustering (Ye, Gao, & Zeng, 2003) in digital image processing, Kim, Bae and Lee (2007) developed a computer system that can rate color-related elements such as the number and types of colors used, the number of color blends, the size of solid-colored areas, and the length of edges where one color changes to another color. The overall variety of color found in a given drawing is a function of these elements. Human knowledge, experience, feeling, and intuition can now be implemented by the knowledge base of an expert system, which is a developing field of artificial intelligence (Giarratano & Riley, 2005).

Expert systems have been proposed as an approach to help find solutions to the problems encountered in art psychotherapy. Kim, Ryu, Hwang, and Kim (2006) recently developed one such expert system that is capable of processing drawing characteristics, psychological symptoms, individual environments, and psychological disorders. Their system is expected to make significant progress in systematizing the knowledge of art psychotherapy. Kim, Kim, Lee, Lee, and Yoo (2006) have improved the above system by increasing its capabilities of consistency maintenance, reliability evaluation, and machine learning. Kim, Yoo, Kim, and Lee (2007) have presented a framework for the expert system knowledge base in art therapy. The expert systems developed by the above authors and others can determine the main, subsidiary, and background colors in a drawing (Kim, 2008) and the types of imbalances in the placement of a drawing on paper (Kim, Kang, & Kim, 2008).

Method

Our premise was that the computer system's ranking of the variety of color in drawings under consideration could be an objective rating tool, and thus aid human raters in art assessments. In our study, the human raters were blinded toward each other's ratings and asked to rank order a sample of drawings by comprehensively comparing the variety of color present in each drawing. The Spearman Rank Correlation Coefficient (RCC) (Walpole & Myer, 2006) was then used to measure interrater reliability. Next, the computer system ranked the same sample of drawings by detecting the number of colors used in each one. In the case of a tie, the drawing with longer edges was given a higher rank. To detect the number of colors used and the length of color edges, we used the computer technologies of color recognition and edge detection as developed by Kim, Bae, and Lee (2007). Finally, correlation between the ranking by human raters and the ranking by the computer system was examined.

We collected a sample of crayon drawings ($N = 52$) by third-, fourth-, and fifth-grade elementary school students with no known history of emotional disorders. We selected

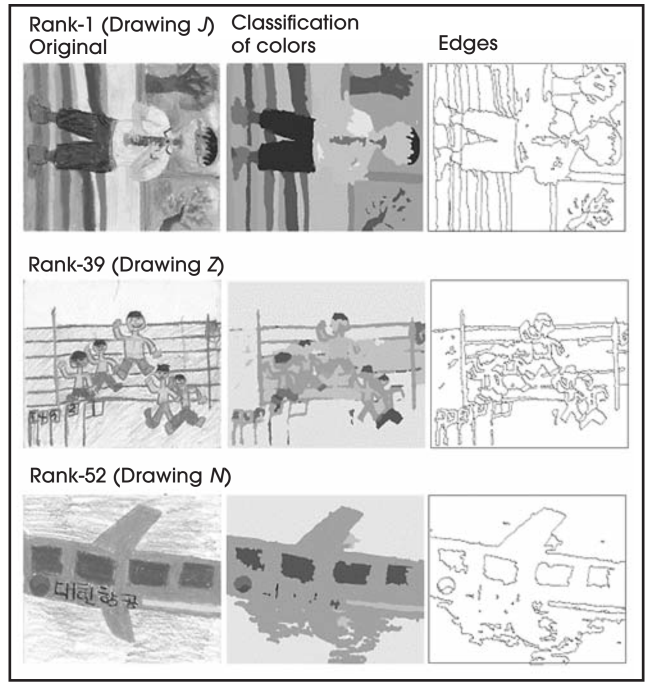

Figure 1 Drawings With Rater-1 Assigned Ranks of Rank-1, Rank-39, and Rank-52

crayons as the medium because of their popularity in Korea. Two art therapy experts compared the variety of color between two drawings and assigned a higher rank to the drawing with greater variety of color. One rater was a registered expressive arts psychotherapist with the Korean Expressive Arts Psychotherapy Association and the other was a color psychology instructor at the Heart & Color School in Korea. The raters were asked to compare the variety of color in each pair of drawings based on their overall impression. We expected that the decision of each rater would inevitably be subjective when rating color because of each individual rater's particular intuition and color perception. Although we asked two raters to rank the variety of color in 52 samples, we did not provide any definitions. We simply asked them to "please rank the variety of color." Had we given them a clearer, more specific definition of variety of color, we might have obtained greater consistency in their ranking, because different raters necessarily have different concepts or understandings of the term "variety."

We assume that the following principle of transitivity applies to variety of color: If drawing A is ranked higher than drawing B, and drawing B is ranked higher than drawing C, then drawing A is ranked higher than drawing C. Thus, all possible pairs need not be compared. The drawing with the greatest variety of color was the most highly ranked, the drawing with the second greatest variety of color is ranked second, and so on. The drawings were rank ordered from 1 to 52; no tie was allowed. In our data collection, we denoted the 52 drawings as *A, B,…Z, a, b,…z*; the two raters as Rater-1 and Rater-2; and the 52 ranks as Rank-1, Rank-2,…Rank-52. Figure 1 shows the sample of the 3 drawings rated by Rater-1 as Rank-1, Rank-39, and Rank-52.

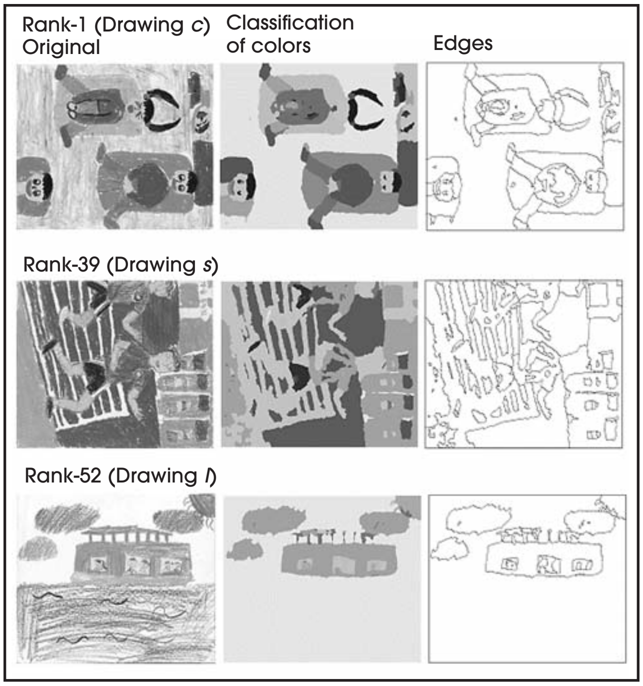

Figure 2 Drawings With Computer Assigned Ranks Corresponding to the Ranks of Rater-1 in Figure 1

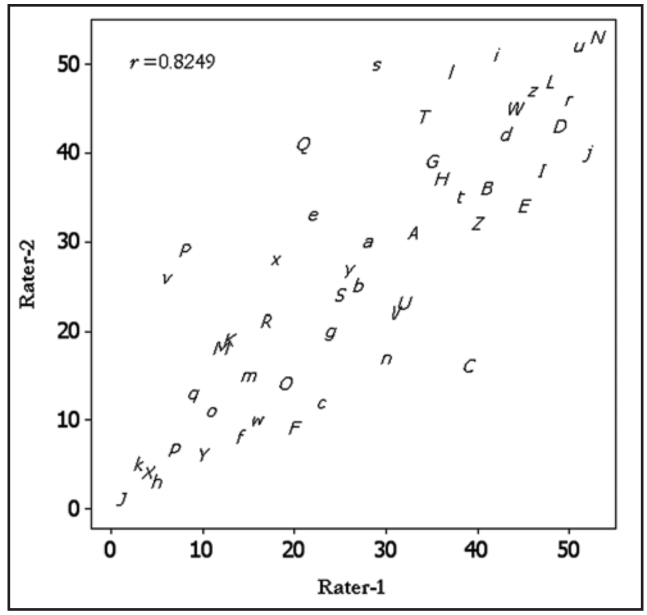

Figure 3
Scatter Plot of the Ranks Assigned by Rater-1 and Rater-2

The computer system rated the variety of color in the same 52 drawings. When 2 drawings were found to have the same number of colors, the one with the longest color edges was ranked the highest. For detecting the number of colors used and the length of color edges, we used the methods developed by Kim, Bae and Lee (2007): after blurring and clustering, the color of each pixel was classified as the closest to one of 47 standard colors defined by the Korean Industry Standard. Blurring and clustering are methods to remove *noise* in crayon drawings. Noise is a technical term in digital image processing that refers to unintended touches due to the thickness of the crayon head. We expressed colors in Munsell's color system, called *HVC* (representing the three elements of colors, hue *H*, brightness *V*, and chroma *C*), which has been accepted as being the most similar to humans' perception of colors (Wan & Kuo, 1998). A specific color is designated by the numerical values of *H, V,* and *C*. For example, $H = 7.5$, $V = 4$, and $C = 14$ is red, $H = 5.8$, $V = 5$, and $C = 12$ is yellow, and $H = 10$, $V = 4$, and $C = 10$ is blue by the Korean Industry Standard. As a measure of similarity between two colors, the distance between them is defined so that the computer can determine the closest color by *HVC* standards. The piece of paper is also divided into pixels, each of which is the final element to be analyzed. For example, if the vertical and horizontal sides of a piece of paper are divided into 480 and 640 points, respectively, then the paper consists of a total of 480 x 640 = 307,200 pixels. Figure 2 shows the drawings rated by the computer system as Rank-1, Rank-39, and Rank-52, for comparison with the human rater's rankings shown in Figure 1.

The RCC was used as a non-parametric measure of correlations between the results given by the two human raters and the computer system. Here, the pairs refer to pairs between Rater-1 and Rater-2, Rater-1 and the computer system, and Rater-2 and the computer system. When the difference between two sets of rankings in the *i*th drawing ($i = 1, 2, \ldots, N$) is represented as d_i, then the RCC is

$$r = 1 - 6 \sum_{i=1}^{N} d_i^2 \,/\, [N (N^2 - 1)].$$

Here, $-1 \le r \le 1$, $r = 1$ is found when the two rankings are identical for every drawing and $r = -1$ when the ranks are completely in the reverse order.

Next, we used the test statistic $z = r\,(N - 1)^{1/2}$ which is asymptotically standard normal as the sample size increases to over 30 (Walpole & Myer, 2006), to determine whether correlations existed between the rankings given by each of the three raters (Rater-1, Rater-2, and the computer system) by testing the null hypothesis of no correlation against the alternative hypothesis of positive correlation. For $N = 52$, the critical region at significance level of 0.05 is $z \ge 1.6448$. If the null hypothesis is rejected, we may conclude that the computer system is able to rate artworks accurately and could thus provide an aid to human raters.

Measure and Test of the Correlation Between the Two Raters

The RCCs in pairs of rankings given by the two human raters and the computer system are presented with the raw data in Table 1. A scatter plot of the rankings made by the two human raters is shown in Figure 3. Table 2 provides a comparison of the drawings that were designated by Rater-1, Rater-2, and the computer system as Rank-1, Rank-13, Rank-26, Rank-29, and Rank-52, using the data collected by the computer system of the number of colors and the length of the color edges in each drawing. Rater-1 ranked drawings *J, f, b, Z,* and *N* respectively Rank-1, Rank-13, Rank-26, Rank-39, and Rank-52 respectively. The number of colors used in each drawing was 25, 20, 23, 19, and 11, respectively. One thus observes that there was a general tendency for the rank order assigned by Rater-1 to be proportional to the number of colors used in each drawing; that is, the greater number of colors in a drawing, the higher rank. The same result occurred with Rater-2. As can be seen in Figure 3, 6 drawings out of a total 52 were designated with exactly the same ranks by the two human raters, *J* of Rank-1, *X* of Rank-3, *o* of Rank-10, *m* of Rank-14, and *N* of Rank-52. As an example of a relatively large difference in ranks, *C* was assigned Rank-38 by Rater-1 and Rank-15 by Rater-2. The RCC and the test statistic are $r = 0.8249$ as in Table 1, $z = 5.8909$ (p-value < 0.0001), respectively. The RCC value shows the high reliability between the two human raters. The large test statistic value also leads to the conclusion that correlation exists between the rankings of these two raters. It is interesting that the raters thought there would be little correlation between their rankings, due to the fact that they were told only to compare "the variety of color" without being given any criteria or definition. Despite this ambiguity, the actual high reliability may be ascribed to a human being's perceptual ability to make collective and comprehensive decisions.

Measure of the Correlation Between the Raters and the Computer

Our next step was to analyze the correlation between the rankings by the computer system and the rankings by each rater. The left and right scatter plots in Figure 4 show the correlations between the rankings by the computer system and Rater-1 and Rater-2, respectively. In the left scatter plot, drawings *R, c, b, U,* and *d* are examples of drawings ranked higher by the computer system than by Rater-1, and drawings *o, Q,* and *g* are examples ranked in the reverse order. Drawing *R* in the left scatter plot is an example of a large difference between ranks: Rank-16 by Rater-1 and Rank-2 by the computer. Figure 5 shows this drawing and compares it with a drawing that was given Rank-16 by the computer system, and one that was given Rank-16 by Rater-2, and another that was assigned Rank-2 by Rater-1. After careful examination, we concluded that in some cases the computer's rating is more appropriate, whereas in other cases the opposite is true.

The RCC between the computer system and Rater-1 ($r = 0.7592$) was found to be higher than between the computer system and Rater-2 ($r = 0.6621$) in Table 1. The correlation of the computer system's rating is relatively high with both raters. Also, we may conclude that high correlations exist between the computer's rating and those by two raters from the values of the test statistics for Rater-1 ($z = 5.4217$ [p-value < 0.0001]) and Rater-2 ($z = 4.7283$ [p-value < 0.0001]). The high reliability between the human raters and the high correlations between the human raters and the computer system validate the usability and usefulness of the computer rating system as an objective aid to the decisions made by human raters.

Discussion

Relatively large differences in the ranks by the two human raters can be seen in the cases of drawings *C, s, Q, P,* and *v*. By eliciting the reasons for this from the two raters, incorporating them in the knowledge base of an expert system, and defining the variety of color in more specific terms, we may reduce the differences between raters. There are relatively large differences between Rater-1 and the computer system in rating the drawings *W, d, b, U, R,* and *c*. Consideration of brightness (*V*) in the *HVC* color space and the list of colors used with the number of clusters may reduce the differences in the ranks given by the raters and the computer system. As an alternative method for rating the variety of color, a regression analysis (Kutner,

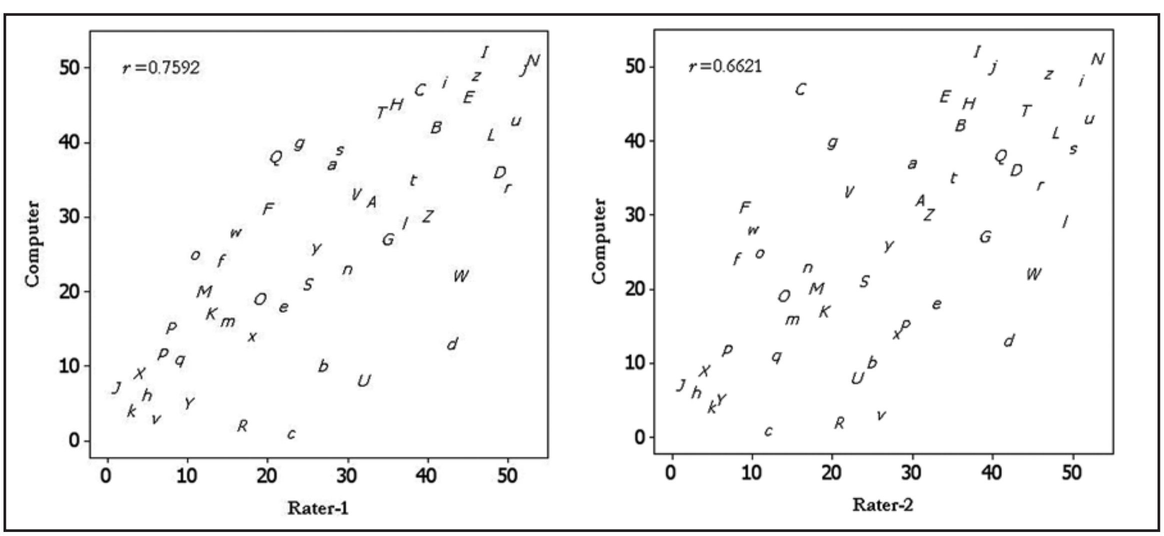

Figure 4
Scatter Plots of the Ranks Assigned by the Computer and Rater-1 (Left) and Rater-2 (Right)

Nachtsheim, Neter, & Li, 2005) could be applied where the dependent variable represents the ranks assigned by the raters and independent variables represent the number of colors, length of color edges, brightness, number of clusters, and specific hues.

The samples were collected from pictures freely drawn by children who had no known history of emotional disorders and who were attending art education classes, not art therapy sessions. If the samples were collected from children with some kind of emotional disorder, or if the samples were a series of drawings made by the same person, we would expect higher reliability and validity. The children in our study used their own 12, 18, 24, or 36 color crayon sets, with the 18-color palette most typically used. In this study we only examined the variety of color in the sample and did not consider the influence of the number of colors in the

Table 1
Rank Correlation Coefficients between Pairs of Ranks by the Two Raters and the Computer with Raw Data

Raters	1	2	3
Drawings (*N* = 52)			
1. Rater-1	❑	.8249	.7592
2. Rater-2		❑	.6621
3. Computer			❑

Drawing Code	Number of colors	Length of edges	Ranks: Rater-1	Ranks: Rater-2	Ranks: Computer	Drawing Code	Number of colors	Length of edges	Ranks: Rater-1	Ranks: Rater-2	Ranks: Computer
A	19	9558	32	30	32	*a*	18	11427	27	29	37
B	17	8942	40	35	42	*b*	23	16847	26	24	10
C	14	14207	38	15	47	*c*	29	10895	22	11	1
D	18	11443	48	42	36	*d*	22	12954	42	41	13
E	15	8916	44	33	46	*e*	21	14726	21	32	18
F	19	9804	19	8	31	*f*	20	12626	13	7	24
G	20	11323	34	38	27	*g*	17	11350	23	19	40
H	16	6386	35	36	45	*h*	25	14188	4	2	6
I	11	5972	46	37	52	*i*	14	7274	41	50	48
J	25	14181	1	1	7	*j*	13	8577	51	39	50
K	22	12000	12	18	17	*k*	25	15908	2	4	4
L	17	9630	47	47	41	*l*	19	12311	36	48	29
M	21	12030	11	17	20	*m*	22	12196	14	14	16
N	11	10607	52	52	51	*n*	20	13881	29	16	23
O	21	13221	18	13	19	*o*	20	12575	10	10	25
P	22	12680	7	28	15	*p*	23	12194	6	6	12
Q	18	10629	20	40	38	*q*	23	14997	8	12	11
R	26	15992	16	20	2	*r*	18	20270	49	45	34
S	21	11369	24	23	21	*s*	17	15274	28	49	39
T	16	12362	33	43	44	*t*	18	11713	37	34	35
U	25	11043	31	22	8	*u*	16	13994	50	51	43
V	19	9209	30	21	33	*v*	25	21364	5	25	3
W	21	9242	43	44	22	*w*	19	12718	15	9	28
X	24	16992	3	3	9	*x*	22	12725	17	27	14
Y	25	14850	9	5	5	*y*	20	12304	25	26	26
Z	19	10982	39	31	30	*z*	13	16713	45	46	49

Table 2
Computer Detection of the Number of Colors and the Length of Color Edges in Sample Drawings in Figure 1.

	Rater-1			Rater-2			Computer		
	Drawing Code	Number of colors	Length of edges	Drawing Code	Number of colors	Length of edges	Drawing Code	Number of colors	Length of edges
Rank-1	*J*	25	11918	*J*	25	11918	*c*	29	10895
Rank-13	*f*	22	12370	*O*	21	11369	*d*	27	12000
Rank-26	*b*	20	10356	*y*	20	12626	*y*	20	12626
Rank-39	*Z*	14	12932	*j*	17	8942	*s*	18	11443
Rank-52	*N*	13	6824	*N*	13	6824	*I*	11	5972

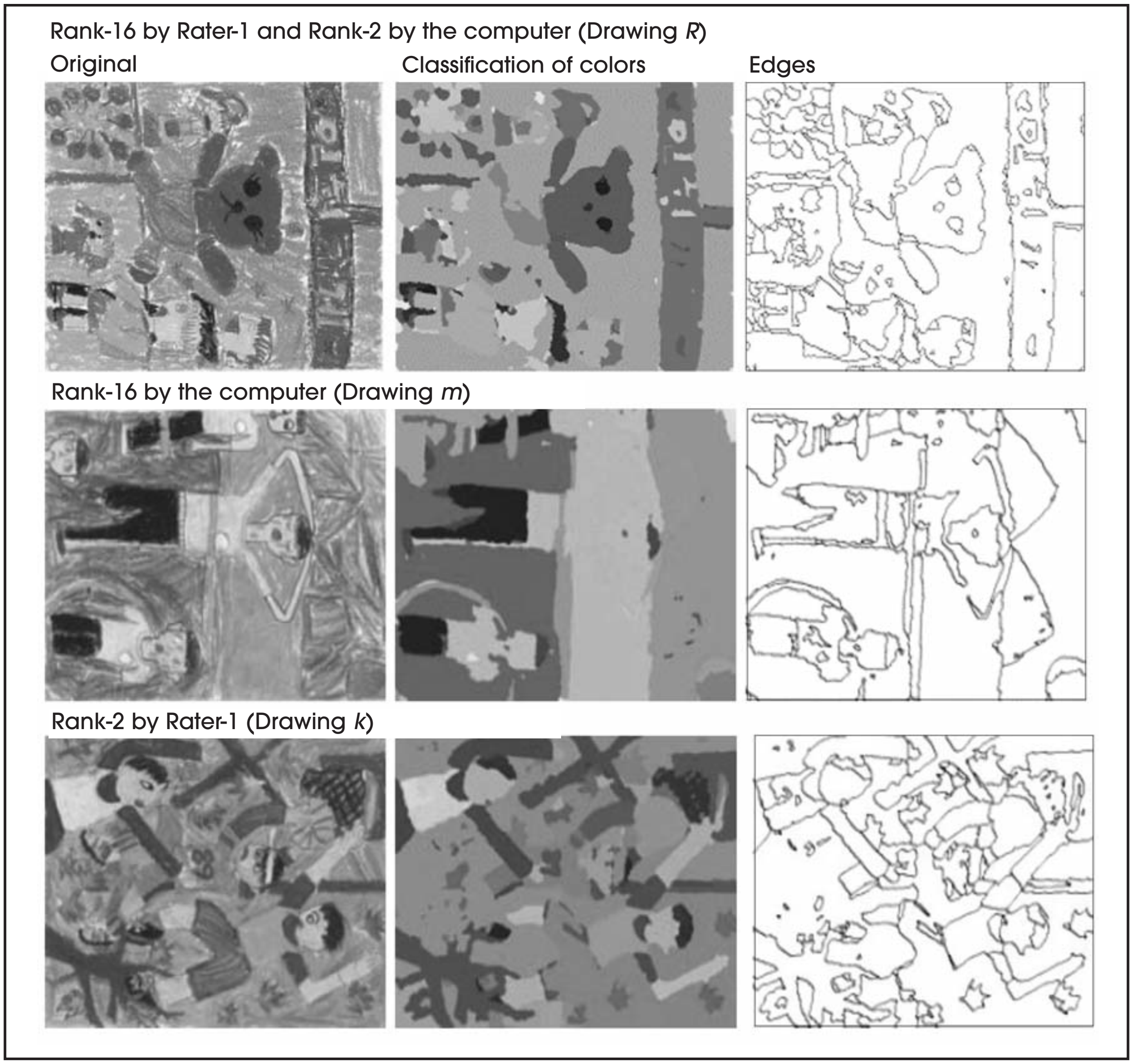

Figure 5 Comparison of Drawings Assigned Rank-16 and Rank-2 by Rater-1 and the Computer

drawing medium on the variety of colors used. It is also worth noting that the samples were collected from classes of typical third-, fourth- and fifth-grade students. We expect that the same conclusions may be found in the drawings of children of other ages with respect to the reliability of human raters and the usability of the computer rating system. However, this can be a subject for a future study.

Conclusion

We have developed a computer system to systematically rate the variety of color in drawings. Such a system has the potential to solve the problem of subjectivity, preconception, and bias in this type of human decision-making. The system evaluates the number of colors used and the length of edges between colors detected through methods drawn from the field of artifical intelligence: blurring and clustering, and technologies of vision. When assessing a sample of 52 children's crayon drawings, the two human raters showed high interrater reliability and the computer system demonstrated high correlation with the human raters. Thus, we can assert that the computer system can serve as a useful aid for human raters. Moreover, the high correlation and the large values of the test statistics suggest the possibility of the computer rating replacing the human decisions, especially in indecisive cases. Furthermore, computer rating has the potential for greater accuracy in the sense that human beings are subject to issues of carelessness or fatigue when rating dozens or hundreds of drawings.

In conclusion, the human raters showed relatively high reliability in their assessment of variety of color. The computer system was verified as an objective rating tool to aid human raters; it could also determine changes in the variety of color across an entire series of drawings. This finding, along with other elements in a drawing, can provide useful information on a patient's diagnosis and emotional status. Computer technologies can contribute to the objectification or quantification of human decisions. This interdisciplinary research of psychology, art psychotherapy, and computer technologies may lead to practical, theoretical, and philosophical progress in the field of art assessments.

References

Betts, D. J. (2006). Art therapy assessments and rating instruments: Do they measure up? *The Arts in Psychotherapy, 33*(5), 422–434.

Cohen, B. M. (Ed.). (1986/1994). *The Diagnostic Drawing Series rating guide.* Unpublished manual.

Francis, D., Kaiser, D., & Deaver, S. P. (2003). Representation of attachment security in the Bird's Nest Drawings of clients with substance abuse disorders. *Art Therapy: Journal of the American Art Therapy Association, 20*(3),125–137.

Gantt, L., & Tabone, C. (1998). The Formal Elements Art Therapy Scale and "Draw a Person Picking an Apple from a Tree." In C. A. Malchiodi (Ed.), *Handbook of art therapy* (pp. 420–427). New York: Guilford Press.

Ghaffurian, M. A. (1995). *Visual arts and healing.* Geelong, Australia: Deakin University.

Giarratano, J., & Riley, G. (2005). *Expert systems: Principles and programming* (5th ed.). Boston: Course Technology.

Gonzalez, R. C., & Woods, R. E. (2002). *Digital image processing* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Gregorian, V. S., Azarian, A., DeMaria, M. B., & McDonald, L. D. (1996). Colors of disaster: The psychology of the "black sun." *The Arts in Psychotherapy, 23*(1), 1–14.

Kim, S. I. (2008). Computer judgment of main color in a drawing for art psychotherapy assessment. *The Arts in Psychotherapy, 35*(2), 140–150.

Kim, S. I., Bae, J., & Lee, Y. (2007). A computer system to rate color-related formal elements in art therapy assessments. *The Arts in Psychotherapy, 34*(3), 223–237.

Kim, S. I., Kang, H. S., & Kim, K. E. (2008). Computer determination of placement in a drawing for art therapy assessments. *The Arts in Psychotherapy, 35*(1), 49–59.

Kim, S. I., Kim, K. E., Lee, Y., Lee, S. K., & Yoo, S. (2006). How to make a machine think in art psychotherapy: An expert system's reasoning process. *The Arts in Psychotherapy, 33*(5), 383–394.

Kim, S. I., Ryu, H. J., Hwang, J. O., & Kim, M. S. H. (2006). An expert system approach to art psychotherapy. *The Arts in Psychotherapy, 33*(1), 59–75.

Kim, S. I., Yoo. S., & Kim, K. E. (2007). A framework for expert system knowledge base in art psychotherapy. In A. R. Tyler (Ed.), *Expert systems research trends* (pp. 181–206). Haupauge, NY: Nova Science.

Kutner, M. H., Nachtsheim, C. J., Neter, N., & Li, W. (2005). *Applied linear statistical models* (5th ed.). New York: McGraw-Hill.

Malchiodi, C. A. (1990). *Breaking the silence: Art therapy with children from violent homes.* New York: Brunner/Mazel.

Oster, G. D., & Gould, P. (2004). *Using drawings in assessment and therapy: A guide for mental health professionals* (2nd ed.). New York: Brunner-Routledge.

Rorschach, H. (1951). *Psychodiagnostics* (5th ed.). Bern, Switzerland: Hans Huber.

Wadeson, H. (1980). *Art psychotherapy.* New York: John Wiley & Sons.

Walpole, R. E., & Myers, R. H. (2006). *Probability and statistics for engineers and scientists* (8th ed.). New York: Macmillan.

Wan, X., & Kuo, C. C. (1998). A new approach to image retrieval with hierarchical color clustering. *IEEE Transactions on Circuits, Systems and Video Technologies, 8,* 628–643.

White, C. R., Wallace, J., & Huffman, L. C. (2004). Use of drawings to identify impairment among students with emotional and behavioral disorders: An exploratory study. *Art Therapy: Journal of the American Art Therapy Association, 21*(4), 210–218.

Ye, Q., Gao, W., & Zeng, W. (2003). Color image segmentation using density-based clustering. *IEEE International Conference on Acoustics, Speech & Signal Processing, 2,* 401–404.