

Measuring icon complexity: An automated analysis

ALEX FORSYTHE, NOEL SHEEHY, and MARTIN SAWEY
Queen's University, Belfast, Northern Ireland

Measures of icon designs rely heavily on surveys of the perceptions of population samples. Thus, measuring the extent to which changes in the structure of an icon will alter its perceived complexity can be costly and slow. An automated system capable of producing reliable estimates of perceived complexity could reduce development costs and time. Measures of icon complexity developed by Garcia, Badre, and Stasko (1994) and McDougall, Curry, and de Bruijn (1999) were correlated with six icon properties measured using Matlab (MathWorks, 2001) software, which uses image-processing techniques to measure icon properties. The six icon properties measured were icon foreground, the number of objects in an icon, the number of holes in those objects, and two calculations of icon edges and homogeneity in icon structure. The strongest correlates with human judgments of perceived icon complexity (McDougall et al., 1999) were structural variability ($r_s = .65$) and edge information ($r_s = .64$).

Small pictorial images, known as icons or symbols, are ever-present in our society; they are used anywhere that language can be used to facilitate the communication of information. Although the terms *icon* and *symbol* are often used interchangeably, there are some subtle differences. Icons often have a physical connection to a particular target or function, whereas symbols are more arbitrary in nature and often have an indirect relationship to their referents (Fenk, 1998; Horton, 1994). Yet, the use of the word *icon* to describe symbols—particularly on computer screens—is so pervasive that it is probably irreversible (Horton, 1994).

There are distinct advantages to the use of pictorial images for communication. Pellegrino, Rosinski, Chiesi, and Siegel (1977) reported that when two objects belong to the same category, pictures depicting them are processed faster than words. Information is also conveyed more directly through the use of images; pictorial road signs are understood easily and accurately and are learned more easily than word signs (Walker, Nicolay, & Stearns, 1965). There is also strong evidence that symbols and icons are more easily identified and better remembered. In tasks involving recall, participants have been able to correctly identify 85%–95% of a set of previously presented graphical symbols, whereas recall for words was significantly lower, particularly in larger learning sets (Standing, 1973). A similar advantage was reported by Muter and Mayson (1986), who found that error rates were halved when users were presented with symbols rather than with text; more

recently, Giannini, Giannini, and Condon (2000) found that the presentation of symbolic slides during lectures facilitated improvements in the test scores of students.

It may be that the memory capacity for pictures is almost limitless (Harber & Hersenson, 1973; Standing, 1973), but the value of an icon or symbol depends largely on the effort required for an accurate interpretation of its meaning. The possibility of creating an unlimited number of icons means that the potential for communication failure is high. In attempting to address this problem, the International Organization for Standardization (ISO) has set a minimum recognition rate of 66.7% (Lindgaard, Chessari, & Ihsen, 1987; Piamonet, Abeysekera, & Ohlsson, 2001), but reports of recognition rates as low as 29% point to difficulties in meeting the ISO criterion (Davies, Haines, Norris, & Wilson, 1998; Piamonet et al., 2001).

The development of design guidelines, informed by contributions from several disciplines, including engineering, ergonomics, and psychology, represents another approach to the development of effective icons. There have been comprehensive reviews of the principal ideas underpinning icon theory (Easterby & Zwaga, 1984; Horton, 1994; Maguire, 1985) and ISO publishes a broad spectrum of recommendations regarding the location of icons, including those pertaining to road vehicles (ISO, 1995) and information technology (ISO, 2000a, 2000b, 2000c, 2000d).

Research into good icon design and usability has been impeded by difficulties in measuring and controlling symbol characteristics. Traditional methods involve the generation of large numbers of picture sets (Alario & Ferrand, 1999; Rogers & Osborne, 1987; Sanfeliu & Fernandez, 1996; Snodgrass & Vanderwart, 1980). Recently, McDougall, Curry, and de Bruijn (1999) used this method to develop the first set of normative values for icons. This

Supplementary material, including an appendix to this article, is available at <http://www.psych.qub.ac.uk/research/projects/aia/>. Correspondence concerning this article should be addressed to A. Forsythe, School of Psychology, Queen's University Belfast, Belfast, Northern Ireland BT7 1NN (e-mail: a.forsythe@qub.ac.uk).

and a subsequent study (McDougall, de Bruijn, & Curry, 2000) have provided a common framework for the evaluation of icon characteristics.

McDougall et al. (1999; McDougal et al., 2000) found three characteristics to be of primary importance in the measurement of symbols and icons: concreteness, distinctiveness, and complexity. The degree to which an icon could be considered concrete or similar to its real-world counterpart was found to be important to the initial understanding of the icon. A more dynamic quality of icons—their distinctiveness—can help users identify and respond to the messages they convey. Distinctiveness cannot be assessed in isolation; it is contingent on the nature of the visual display in which an icon is located. The third characteristic, icon complexity, is reported to be important in search efficacy. The amount of detail or intricacy within an icon influences the rate at which it can be detected, with very simple or very abstract icons being detected faster. Interestingly, icon complexity was found to be unrelated to how concretely similar an icon was to its referent (McDougall et al., 1999; McDougall et al., 2000).

Making major alterations to an icon—such as rescaling or inverting its black/white contrasts—will almost certainly change the way the icon is interpreted. To what extent can designers change the structural properties of icons without changing behavioral responses to them? Currently, there are relatively few methods available to predict behavioral changes from structural alterations.

One technique involves the iterative testing of modified icons using samples drawn from populations of potential users. User familiarity means, however, that different samples are required at each stage of the modification process, and the designer is left to guess when it might be appropriate to engage in the costly and laborious process of normative reassessment. Likewise, in the development of an entirely new icon set, the gathering of subjective judgments from users should occur throughout the design process, particularly at the beginning and end of set development (McDougall et al., 2000). An automated system would enable designers to check iteratively throughout the design process, without the necessity of recruiting large numbers of human raters.

As a decision-making aid, an automated system would make the process of icon development and modification a less speculative, more cost-effective activity.

Why Should Automated Measurement Be Possible?

The characteristics of icon concreteness and distinctiveness are, to some degree, dynamic properties, in the sense that the degree to which an icon is judged to be a concrete representation is determined partly by an observer's familiarity with the target object or function. The distinctiveness of an icon will depend not only on its structural features, but also on the visual landscape in which it is situated (McDougall et al., 2000). Judgments of icon complexity, however, may be somewhat less sensitive to factors of user familiarity and visual context and may be computation-

ally tractable through the use of an automated measurement procedure.

McDougall et al. (1999) has reported on the success of one semi-automated system of icon measurement. Garcia, Badre, and Stasko (1994) developed a technique based on a calculation of several icon features, including the number of closed and open figures and horizontal and vertical lines. This metric was developed primarily as a measure of icon concreteness, and, using this measure, Garcia et al. reported that icons that are pictorially similar to their real-world counterparts are more likely to be judged as complex. Although there is some disagreement as to whether icon complexity and icon concreteness are related constructs (McDougall et al., 1999), the high correlations between the metric of Garcia et al. and the McDougall et al. (1999) measures of icon complexity ($r_s = .73$) suggest that the Garcia et al. measure of complexity is able to measure the amount of detail in an icon in a way similar to that of human observers. However, Garcia et al.'s techniques are painstaking for all but the simplest icons. In the study reported here, it was considered whether a more automated system could be used to quickly produce reliable estimates of perceived icon complexity.

Several theories point to the possibility of developing a more computationally driven approach. First, it is generally accepted that the degree to which an icon is judged to be concrete or similar to its target object is important in providing meaning to the user (Guastello & Traut, 1989; McDougall et al., 2000; Mead & Modley, 1968; Schwartz & Phillippe, 1991). It has been reported that there is not necessarily a direct relationship between the icon dimensions of concreteness and complexity (McDougall et al., 1999), but icons regarded as similar to their targets do tend to be more detailed and explicit in nature (Arend, Muthig, & Wandmacher, 1987; Rogers & Osborne, 1987). This extra detail is of particular importance to the novice user (see, e.g., Maguire, 1985; McDougall et al., 2000; Mead & Modley, 1968; Raeder, 1985); for example, essential elements, such as the wings on an airplane or the handle on a suitcase, give more explicit meaning to an icon, thereby facilitating faster identification (Bruyas, Le Breton, & Pautie, 1998). Other advantages of using icons that are highly detailed in nature include the ability of designers to convey action information, such as when a physical object requires manipulation (Figure 1) or in attracting the eye to small details that can aid in recognition, such as the luggage label in Figure 2A.

This extra detail however, comes at a cost. Regardless of experience, more abstract icons—which are not necessarily good examples of a prototype—may elicit faster responses than icons offering more concrete representations (Arend et al., 1987; McDougall et al., 2000). One explanation for this is that the observer seeks a functional explanation for any additional icon objects; hence, processing time increases (Bruyas et al., 1998; Nielson, 1993). Consequently, it takes longer to process the suitcase in Figure 2A than that in Figure 2B. A further design feature that can add to the complexity of an icon is the use of



Figure 1. Count icon.

color. Color is useful in attracting the attention of the user and for organizing and grouping related icon concepts; however, general advice is that, where possible, the use of color should be avoided. At best, color offers no advantage to user performance (Christ, 1977; Tullis, 1981), and at worst it reduces the speed at which information is processed (Biederman & Ju, 1988; Ostergaard & Davidoff, 1985). Icons should be designed in black and white, with color added only for improvement (Horton, 1994).

McDougall et al. (1999; McDougall et al., 2000) have reported that the level of complexity within an icon bears no direct relationship to the degree to which it can be considered concrete or representative of its target. Concreteness is related to the meaning implied by an icon, whereas complexity is related to recognition time. One explanation for this is focused on the impact of including additional elements in an icon (Horton, 1994; Nielsen, 1993). Many icons in common use are intended to convey information about their targets and the operation to be applied to them. For example, an icon depicting a file held within a clamp can be used to refer to a generic target (any file stored on a computer) and the operation of compression. Depicting both an object and an operation usually involves the inclusion of more elements to clearly communicate the intended meaning. Additional attentional processes are involved in building a cognitive representation of an icon, because of the need to integrate the greater number of constituent elements. Simple image properties are extracted from an image in parallel, and these properties are then combined to form objects of a particular shape, color, and size (see, e.g., Treisman, 1986; Treisman & Gelade, 1980; Treisman & Souther, 1985).

An alternative explanation that lends itself to measurement consists in the Gestalt principles of figure/ground and minimum. These center on the way in which higher order grouping processes and selective attention guide the ordering and recognition of image features. For example, the law of simplicity suggests that, when viewing a complex image, the human perceptual system operates selectively to secure the most economical form (Hochberg, 1968; Hochberg & Brooks, 1960; Neisser & Becklen, 1975). Hochberg even went so far as to develop a semi-automated measure of image complexity, arguing that relying solely on human judgments of what makes an image "complex" would mean that we have no way of predicting just how advanced or simple an image will actually appear. His calculations demonstrated that it was possible to predict how viewers would "see" an image: The more in-

terior angles, different angles, and lines an image had, the more likely it was to be perceived in three dimensions.

Gestalt grouping processes are also considered to be important in the perception of form; the largest area of an image will always be considered background, with contour and shape considered foreground (Hochberg & Brooks, 1960). Perhaps this explains why such grouping processes are important in the perception of image complexity. For example, a global-feature superiority effect has been reported in the high-speed selection of icons, with responses to features such as size and shape being faster than those to the structural detail of an image (Arend et al., 1987; Wandmacher & Arend, 1985). Psychophysical evidence suggesting that higher order grouping processes operate in this way can be found in studies showing that the brain registers variations in an image as changes in intensity, and it is these coarse and fine changes that provide detail and local information about a stimulus (Beck, Graham, & Sutter, 1991; Harwerth & Levi, 1978; Sutter, Beck, & Graham, 1989; Vassilev & Mitov, 1976). Coarse scales are thought to be treated by the brain as low-frequency components and are processed faster than high-frequency components obtained from local information. This difference in processing speed would seem to be a function of image complexity: When an object is of a detailed nature, its global attributes are processed much faster than its local ones (Hoeger, 1997; Parker, Lishman, & Hughes, 1997).

Although the theoretical positions underpinning feature integration theory, Gestalt explanations, and psychophysics research are all quite different, they all suggest that the measurement of icon complexity may be computationally tractable. If this is so, then it should be possible to develop an automated system that would inform designers about human estimations of users' perceptions of icon complexity. Such a system could prove to be a useful decision-making aid to designers who wish to create an entirely new icon set or make alterations to the properties of an existing one.

Method

Matlab (MathWorks, 2001) is an integrated commercial package with powerful mathematical algorithms and visualization utilities for the acquisition, analysis, and exploration of data. With its Image Processing Toolbox, Matlab can be used to recover detailed structural information from a black and white, grayscale, or color image.

The Toolbox has numerous image-processing functions and, on the basis of the theoretical explanations of icon

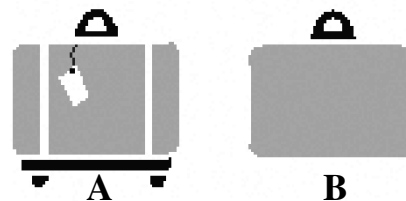


Figure 2. Luggage icon.

complexity reviewed above, six of these functions were applied to a random sample ($n = 68$) of the McDougall et al. (1999) icons. The function measures were correlated with metrics developed by Garcia et al. (1994) and McDougall et al. (1999). High correlations would indicate similar measurements of icon complexity.

The Gestalt argument (Arend et al., 1987) that distinct global features are important to perceptual speed was tested first. It was considered that the most global characteristic of an icon would be its figural form, which was calculated using a measure of icon foreground. High correlations between this measure and the metrics of Garcia et al. (1994) and McDougall et al. (1999) would indicate that grouping principles are important factors in human ratings of complexity.

When preparing the McDougall et al. (1999) icon set for processing, Matlab treats a black and white, or binary, icon as an array of 1s and 0s. This is essentially a data file in which the image is stored as a matrix of pixels. On white paper, black normally (but not always) represents the foreground, and white represents the background. Matlab, on the other hand, considers white to be an “on” pixel, giving it the value 1, and black to be an “off” pixel, giving it the value 0. Thus, before any analysis was carried out, the representation of all the icons in Matlab was reversed, with 1s becoming 0s and vice versa.

Icon foreground is defined as the area taken up by pixels that are switched on (i.e., black on paper, but with the value of 1 in the Matlab representation of the icon). Mat-

lab assigns diagonally connected pixels extra weighting. This is because the discreteness necessarily introduced by the digital representation of an image means that a given number of pixels arranged diagonally represents a longer line than the same number of pixels arranged horizontally or vertically.

A processing convention was then established to define connections between pixels and, hence, an icon’s component parts. All pixels touching both horizontally and vertically (but not diagonally) were considered to be connected. This is referred to by Matlab as a 4-connected neighborhood, and the difference between this and an 8-connected neighborhood can be observed in panels B and C of Figure 3. Essentially, a 4-connected neighborhood produces a finer image.

Four additional calculations were performed using this convention. The feature integration of icon elements (Bruyas et al., 1998; Nielson, 1993; Treisman, 1986) was tested by counting the number of discrete objects within each icon, an object being defined as a set of 4-connected pixels. Figure 4 shows an example of the objects count; the separate objects are shown in various shades of gray, and this icon received an object count of 14. The number of holes within those objects was also counted by calculating the Euler number of the image. A negative Euler number indicates that the number of holes in the image is greater than the number of objects. It was considered that an icon rated as complex would contain more elements than an abstract icon and that these elements would them-

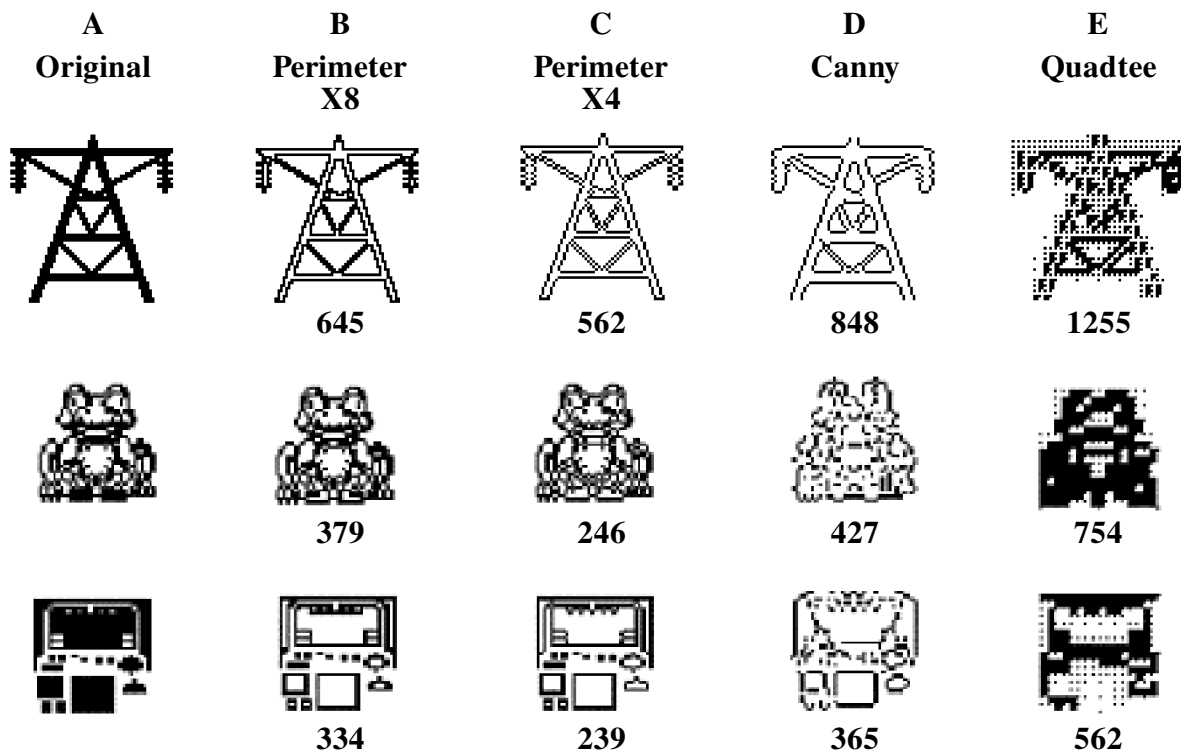


Figure 3. Automated measures morphs with metric scores.



Figure 4. Hierarchy icon.

selves be complex in nature, having more local detail (i.e., holes).

To test the importance of changes in image structure in the perception of icon complexity, two perimeter-detection measures and one structural analysis was performed. If structural variability were an important factor in the perception of icon complexity, it would be indicated by high correlations between these measures and the metrics of Garcia et al. (1994) and McDougall et al. (1999). Examples of how these various metrics affect an icon’s appearance can be observed in Figure 3.

Perimeter-detection measures work by examining the changes in intensity that occur at the edges of an image. Edges are located using two criteria that examine areas in the icon where there is a rapid change in image intensity. Either a change in intensity must be larger than a predetermined threshold (perimeter detection provides a number of estimators that can be used to specify sensitivity) or an edge will be detected where the intensity derivative has a zero crossing. For the purposes of this study, zero crossings occur at the places where negative and positive pixels are adjacent. For a pixel to be considered an edge pixel, it must be activated (on) and connected to at least one non-activated (off) pixel. This is a simplified version of more general detectors, such as Canny and Sobel, which calculate the gradient of intensity values for close-by pixels in color or grayscale images.

A limitation of this analysis is that thicker lines are awarded higher scores than thinner ones. This is because the perimeter calculation rates a thick line as having two edges rather than one. The selection of a 4-connected neighborhood compensates for this problem to some degree, since it produces a finer image (see Figure 3C). However, an edge that might be blurred or difficult for a user to de-

tect may be included superfluously in a Matlab perimeter-detection calculation.

To allow for both of these considerations, a further analysis was performed using the Canny perimeter-detection calculation. The advantage of the Canny method over other perimeter-detection methods is that it works by using two thresholds to detect strong and weak edges and includes the weak edges in the output only if they are connected to strong edges. This means that truly weak edges will be detected in the analysis, but noise—such as shadow or shading—will be ignored. The second icon in Figure 3D (frog) is a good example of the power of the Canny analysis. Whereas the perimeter measures (Figures 3B and 3C) retain elements of shading and shadow in their analysis, the Canny analysis retrieves a much more skeletal impression. Furthermore, since Canny is able to deal with grayscale and binary images, it is unnecessary to treat the data before processing.

Finally, Quadtree decomposition examined homogeneity in the icons. It works by iteratively subdividing the icon into quarters; each subblock is then tested for homogeneity. The subdivision continues until the resulting subblock is homogeneous; hence, a large number of small, homogeneous blocks indicates a highly structured image (see Figure 3E). A small number of largish blocks indicates a more homogeneous image.

Results

The means, standard deviations, and skew for the calculated icon characteristics are displayed in Table 1. McDougall et al. (1999) recorded two measures of complexity for this icon set, one using the Garcia et al. (1994) metric and one involving subjective ratings. These scores are also included in Table 1.

The distribution of scores for several of the measures was positively skewed at more than twice the standard error. To correct this, a log₁₀ transformation was performed on the Garcia et al. (1994), Foreground, Objects, and Holes metrics. This transformation corrected the distribution of data for these metrics; however, rudimentary analysis of the underlying data for these measures presented some anomalies. Consequently, the corrected distributions were artificially inflated for all but the Garcia et al. measures, and only the remaining ratings and this measure were correlated.

Table 1
Descriptive Statistics for Icon Attributes

Metric	<i>M</i>	<i>SD</i>	Skew	Kurtosis
McDougall et al. (1999) complexity	2.75	.78	-.31	5.8
Garcia et al. (1994)	9.33	9.31	2.05	-.33
Foreground	760.62	593.45	2.59	10.44
Objects	13.17	21.00	3.04	11.83
Holes	2.8	4.6	2.47	6.19
Perimeter	384.08	228.26	.859	-.06
Quadtree	733.82	396.32	.79	-.36
Canny	329.34	181.52	.65	-.79

Table 2
Spearman Correlations Between Icon Attributes

Metric	McDougall et al. (1999)	Garcia et al. (1994)	Perimeter	Quadtree
McDougall et al. (1999) complexity	1.00			
Garcia et al. (1994)	.75*	1.00		
Perimeter	.64*	.66*	1.00	
Quadtree	.65*	.65*	.94*	1.00
Canny	.49*	.60*	.88*	.84*

*Correlations significant at $p < .0008$.

Given the large number of analyses performed, a Bonferroni adjustment was applied and the criterion for statistical significance was set at $p < .0008$. A caveat to this adjustment is that it greatly increases the likelihood of Type II error.

The analysis (Table 2) replicated the moderately strong correlations reported by McDougall et al. (1999) between their ratings of complexity and those of the Garcia et al. (1994) metric. This supports the argument that they may be tapping similar constructs.

The Garcia et al. (1994) and McDougall et al. (1999) metrics both correlated with the structural analysis (Quadtree) score ($r_s = .65$, $p < .0008$) and with the perimeter-detection score ($r_s = .66$ and $r_s = .64$, respectively; $p < .0008$). The Canny correlations with Garcia et al. ($r_s = .60$) and McDougall et al. (1999; $r_s = .49$) are smaller but statistically significant. These findings suggest that image-processing techniques can measure icon characteristics in ways that approximate more time-consuming measures. The correlations also suggest that ratings of icon complexity may be linked to homogeneity in image structure.

Discussion

Positively skewed distributions were observed for several of our measures as well as for the Garcia et al. (1994) metric. Skewed distributions are commonly caused by sampling bias, nonnormal distribution of the characteristics of the items being measured, or the sensitivity of the measurement tool. The first possibility can be discounted, because the sample of icons was randomly drawn from those published by McDougall et al. (1999). The second fails to explain why the McDougall et al. (1999) norms are approximately normally distributed, whereas our measures and those of Garcia et al. are positively skewed: If icon characteristics are not normally distributed, that should be observed consistently.

The third possibility suggests that enumeration methods of the type used by Garcia et al. (1994) and by us yield positively skewed distributions because they are based on the systematic application of rigorous definitions of icon attributes. Stringent enumeration techniques would be expected to yield measures of icons that pile up on the lower end of the scale, whereas the judgments of human observers are probably based on a less meticulous analysis and produce a more normal distribution of scores. This is supported indirectly by evidence indicating that many image-processing measures yield poor estimates of

human judgments of icon complexity. For example, calculations on objects within an icon are based on connections between pixels such that discrete objects will be detected only when there are breaks between the pixels. Thus, because the nodes in Figure 5 are all linked by lines, our computational measures detect only one discrete object. Feature integration theory contends that the human perceptual system is much more sophisticated than this, with the nodes and lines of the hierarchy icon being extracted separately (Treisman & Gelade, 1980). Specifically, when an image is made up of more than one separable feature, individual features are attended to and processed along independent dimensions (Treisman & Gelade, 1980). Thus, greater cognitive effort is involved in the integration of an icon composed of a number of separable elements than in that of an icon that is more integral in nature. According to this view, icon complexity is partly a function of the cognitive effort involved in making an evaluative judgment.

If the problem of object identification was limited to a few icons, this metric might still have proved a useful general estimate of the perceived characteristics of icon properties. However, the problem is quite substantial: Fewer than three objects were detected in almost 50% of the icons we examined. Furthermore, changes to icons that we thought might have led to an improvement in object counting produced disappointing results. For example, the icon in Figure 6 depicts the concept *debug*. The different objects detected by the objects metric are displayed in grayscale. If an 8-connected neighborhood convention is applied (meaning that all touching pixels are considered connected), the icon will receive a count of 30 (Figure 6A). The problem is not rectified by applying a convention that considers only pixels that are connected horizontally and vertically (a 4-connected neighborhood). Although a more conservative count of 2 is obtained from the metric (Figure 6B), the score is awarded for the entire image plus one rogue pixel (shown in gray). Furthermore,

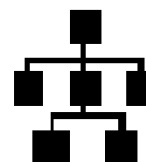


Figure 5. Objects and holes metric.

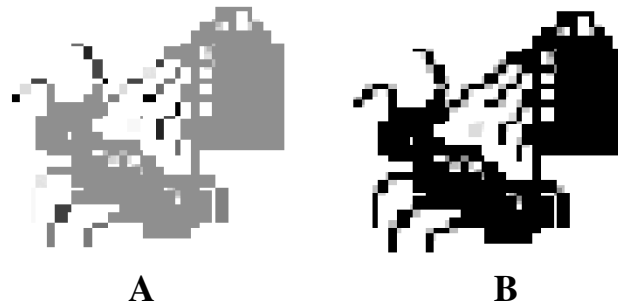


Figure 6. Debug icon.

deletion of either the insect or the repellent from the icon without removing the rogue pixel would still produce an object count of 2. Similar problems were encountered with many of the icons.

The problems posed by this type of analysis arise from the fact that icons consisting of multiple objects tend to have many touching pixels. A similar problem arises when a count of holes is used as an estimate of icon complexity. Image-processing software can detect holes that may be imperceptible to the naked eye, to the extent that even dense icons, such as the one depicted in Figure 5, often received a hole count. In this case, the image-processing software detected a single tiny hole. These examples illustrate that the skewed distribution of data was caused by some fundamental properties of the way the software measured the icons. As such, it could not be treated as a form of measurement error, and, for this reason, a \log_{10} transformation of the raw data was considered inappropriate.

Foreground was chosen as a possible measure of the global features of an icon, but automated measurement also proved difficult. The significantly skewed data suggested that processes unlike those involved in the human perception of figure and ground were in operation. Whereas it may sometimes be possible to treat black areas of an icon as foreground, problems arise when this definition is applied globally. For example, treating black areas as foreground is acceptable when the simple icon in Figure 7A is being processed, but it is unsatisfactory when it is applied to the analysis of the icon in Figure 7B. Although it would be possible to analyze icons on a case-by-case basis, adapting the algorithm accordingly, the procedure would be no

more economical than using human raters for the same purpose.

The foreground metric was even less effective at differentiating icon detail from icon foreground. Detail can be particularly important in providing function cues to the user by drawing the eye to significant areas of information within the icon (Horton, 1994). However, delineating areas that represent icon foreground from those denoting icon detail is very subjective, as is illustrated by the icon in Figure 7C. A simple measure of foreground is too crude a proxy for the complex grouping processes involved in human perception, and it appears likely that this measure is merely a sophisticated calculation of icon area, as represented by pixels that are switched on.

More reliable metrics for estimating human judgments of icon complexity are available in the Perimeter and Canny perimeter-detection measures and the Quadtree structural variability measure. These metrics yield distributions that are much less skewed and that correlate strongly with the complexity ratings reported by Garcia et al. (1994) and McDougall et al. (1999). They also point to the potential importance of structural variability in judgments of icon complexity. As was previously noted, the structural variability in an icon is probably determined by changes in intensity that occur at the edges of an image. Furthermore, the Quadtree analysis of image structure, which also correlates highly with the Garcia et al. and McDougall et al. (1999) measures, is based on changes in image intensity. Highly structured areas correspond to high numbers of small, homogeneous blocks, and these blocks collect around image boundaries. As is the case with the

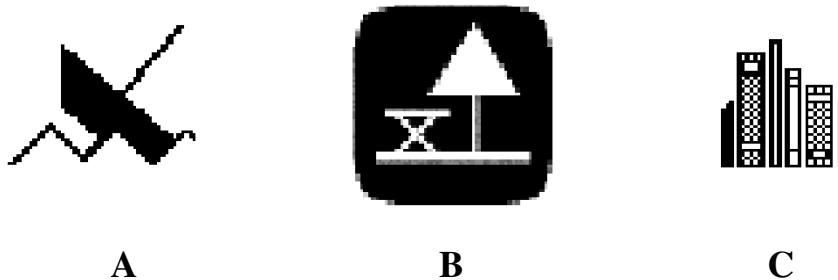


Figure 7. Foreground metric.

Perimeter and Canny measures, these boundaries are areas in which large changes in image intensity take place.

These findings add support to the argument that a complex image is one that has a large number of edges, and it is these edges that slow down the perceptual process (Hoeger, 1997; Parker et al., 1997). Hummel and Biederman (1992) used contour deletion to make a similar point. They argued that perceptual slowdown might be caused by an inability to make a rapid semantic link between the icon and its target. Although semantic processing is not slow in and of itself, the physical properties that bind an image must be processed before semantic access can occur. This would also explain why differences in response rates between complex and abstract icons do not change as a function of learning (Arend et al., 1987; McDougall et al., 1999). Regardless of experience, a more detailed object will take longer to reach semantic processing, because the preprocessing of many edges is perceptually demanding.

Other Relationships

Several of the automated measures correlated significantly with each other. In particular, the Canny analysis correlated highly with both the Perimeter and Quadtree metrics. What difference there is between the measures is almost certainly due to the slightly different ways in which Matlab implements each of the analyses. However, for nonbinary images, such as colored icons and photographs, a Canny analysis could be extremely useful in removing the effects of noise from a calculation.

A preliminary analysis suggests that these automated calculations are reasonably robust, with all calculations remaining stable when the icons are rotated. When the icons are resized, counts of objects and object holes remain stable, but counts of Perimeter, Quadtree, and Foreground areas will tend to vary with the degree of modification.

Complexity and Semantics

Although moderate to strong correlations were observed between the automated measures of icon complexity and human judgments, they were not perfect. The lack of convergence is almost certainly due to the fact that the image-processing algorithms we applied to the icons are poor approximations of the more complex processes embedded in human perceptual judgment. These algorithms do not process semantic information, whereas there is considerable psychophysical evidence that human observers cannot process the structure of an image independently of its semantic content. Boucart and Humphreys (1992) used a sequential form-matching task to manipulate semantic relations between pictures that were physically identical, physically different, semantically related, or semantically unrelated. When the distractor icon was semantically related, reaction times were found to be longer than when it was unrelated. This and subsequent studies (Boucart & Humphreys, 1995, 1997) suggest that the processing of an image form cannot be completed without some semantic analysis. This may also explain

why McDougall et al. (1999) found human judgments of icon familiarity to correlate negatively with their evaluations of icon complexity ($r_s = -.31, p < .001$): As ratings of icon familiarity decrease, complexity tends to increase. Computer-based algorithms should not be so influenced, and to determine whether this is the case, a sample of 70 icons was assembled, which consisted of all those rated as unfamiliar (scoring < 3) in McDougall et al.'s (1999) data set and judged to be very simple by our perimeter measure (scoring < 400). The association between the Perimeter measure of complexity and human judgments of familiarity is close to zero ($r_s = -.05$). Thus, automated measures of the visual complexity of an icon are unaffected by the semantic analyses that human observers cannot suppress when judging icon complexity.

Conclusion

To what extent can designers change the structural properties of icons without changing behavioral responses to them? Decision-making aids based on the image-processing measures reported here can significantly reduce the degree of speculation involved in arriving at an estimate. In practical terms, this means that a designer can modify any of the icons in the set published by McDougall et al. (1999) and, using the Perimeter and Canny perimeter-detection measures together with the Quadtree structural variability measure, can accurately estimate its perceived complexity without incurring the delays and expense associated with a new usability validation study.

REFERENCES

- ALARIO, F.-X., & FERRAND, L. (1999). A set of 400 pictures standardized for French: Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behavior Research Methods, Instruments, & Computers*, **31**, 531-552.
- AREND, U., MUTHIG, K.-P., & WANDMACHER, J. (1987). Evidence for global feature superiority in menu selection by icons. *Behaviour & Information Technology*, **6**, 411-426.
- BECK, J., GRAHAM, N., & SUTTER, A. (1991). Lightness differences and the perceived segregation of regions and populations. *Perception & Psychophysics*, **49**, 257-269.
- BIEDERMAN, I., & JU, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive Psychology*, **20**, 38-64.
- BOUCART, M., & HUMPHREYS, G. W. (1992). Global shape cannot be attended without object identification. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 785-806.
- BOUCART, M., & HUMPHREYS, G. W. (1995). Automatic access to object identity: Attention to global information, not to particular physical dimensions, is important. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 584-601.
- BOUCART, M., & HUMPHREYS, G. W. (1997). Integration of physical and semantic information in object processing. *Perception*, **26**, 1197-1209.
- BRUYAS, M. P., LE BRETON, B., & PAUZIE, A. (1998). Ergonomic guidelines for the design of pictorial information. *International Journal of Industrial Ergonomics*, **21**, 407-413.
- CHRIST, R. E. (1977, October). Four years of color research for visual displays. In *Proceedings of the Human Factors Society 21st Annual Meeting*. San Francisco.
- DAVIES, S., HAINES, H., NORRIS, B., & WILSON, J. R. (1998). Safety pictograms: Are they getting the message across? *Applied Ergonomics*, **21**, 15-23.
- EASTERBY, R., & ZWAGA, H. (1984). *Information design: The design and evaluation of signs and printed material*. New York: Wiley.

- FENK, A. (1998). Symbols and icons in diagrammatic representation. *Pragmatics & Cognition*, **6**, 301-334.
- GARCIA, M., BADRE, A. N., & STASKO, J. T. (1994). Development and validation of icons varying in their abstractness. *Interacting With Computers*, **6**, 191-211.
- GIANNINI, A. J., GIANNINI, J. N., & CONDON, M. (2000). Use of tangential visual symbols to increase the long-term learning processes: Applications of linkage in teaching pharmacological principles of addiction. *Journal of Clinical Pharmacology*, **40**, 708-712.
- GUAZZELLO, S. J., & TRAUT, M. (1989). Verbal versus pictorial representations of objects in a human-computer interface. *International Journal of Man-Machine Studies*, **31**, 99-120.
- HARBER, R. N., & HERSHENSON, M. (1973). *The psychology of visual perception*. New York: Holt, Rinehart & Winston.
- HARWERTH, R. S., & LEVI, D. M. (1978). Reaction time as a measure of suprathreshold grating detection. *Vision Research*, **18**, 1579-1586.
- HOCHBERG, J. E. (1968). *Perception* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- HOCHBERG, J. E., & BROOKS, V. (1960). The psychophysics of form: Reversible perspective drawings of spatial objects. *American Journal of Psychology*, **73**, 337-354.
- HOEGER, R. (1997). Speed of processing and stimulus complexity in low-frequency and high-frequency channels. *Perception*, **26**, 1039-1045.
- HORTON, W. (1994). *The icon book. Visual symbols for computer systems and documentation*. New York: Wiley.
- HUMMEL, J. E., & BIEDERMAN, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review*, **99**, 480-517.
- INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (1995). *Road Vehicles—Symbols for controls, indicators and tell-tales* (ISO 2575). Geneva: Author.
- INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (2000a). *Information technology—User system interfaces and symbols—Icon symbols and functions: Part 1. Icons—General* (ISO/IEC 11581-1). Geneva: Author.
- INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (2000b). *Information technology—User system interfaces and symbols—Icon symbols and functions: Part 2. Object icons* (ISO/IEC 11581-2). Geneva: Author.
- INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (2000c). *Information technology—User system interfaces and symbols—Icon symbols and functions: Part 3. Pointer icons* (ISO/IEC 11581-3). Geneva: Author.
- INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (2000d). *Information technology—User system interfaces and symbols—Icon symbols and functions: Part 6. Action icons* (ISO/IEC 11581-6). Geneva: Author.
- LINDGAARD, G., CHESSARI, J., & IHSEN, E. (1987). Icons in telecommunications: What makes pictorial information comprehensible to the user? *Australian Telecommunication Research*, **21**, 17-29.
- MAGUIRE, M. C. (1985). A review of human factors guidelines and techniques for the design of graphical human-computer interfaces. *Computer & Graphics*, **9**, 221-235.
- MATHWORKS (2001). *Image Processing Toolbox, user's guide Version 3*. Boca Raton, FL: CRC Press.
- MCDougALL, S. J. P., CURRY, M. B., & DE BRUIJN, O. (1999). Measuring symbol and icon characteristics: Norms for concreteness, complexity, meaningfulness, familiarity, and semantic distance for 239 symbols. *Behavior Research Methods, Instruments, & Computers*, **31**, 487-519.
- MCDougALL, S. J. P., DE BRUIJN, O., & CURRY, M. B. (2000). Exploring the effects of icon characteristics on user performance: The role of icon concreteness, complexity and distinctiveness. *Journal of Experimental Psychology: Applied*, **6**, 291-306.
- MEAD, M., & MODLEY, R. (1968). Communications among all people, everywhere. *Natural History*, **77**, 56-63.
- MUTER, P., & MAYSON, C. (1986). The role of graphics in item selection from menus. *Behaviour & Information Technology*, **5**, 89-95.
- NEISSER, U., & BECKLEN, R. (1975). Selective looking: Attending to visually specified events. *Cognitive Psychology*, **7**, 480-494.
- NIELSEN, J. (1993). Noncommand user interfaces. *Communications of the ACM*, **36**, 83-99.
- OSTERGAARD, A. L., & DAVIDOFF, J. B. (1985). Some effects of color on naming and recognition of objects. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 579-587.
- PARKER, D. M., LISHMAN, J. R., & HUGHES, J. (1997). Integration of spatial information in human vision is temporally anisotropic: Evidence from a spatiotemporal discrimination task. *Perception*, **26**, 1169-1180.
- PELLEGRINO, J. W., ROSINSKI, R. R., CHIESI, H. L., & SIEGEL, A. (1977). Picture-word differences in decision latency: An analysis of single and dual memory models. *Memory & Cognition*, **5**, 383-396.
- PIAMONET, D. P. T., ABEYSEKERA, J. D. A., & OHLSSON, K. (2001). Understanding small graphical symbols: A cross-cultural study. *International Journal of Industrial Ergonomics*, **27**, 399-404.
- RAEDER, G. (1985). A survey of current graphical programming techniques. *IEEE Computer*, **18**, 11-25.
- ROGERS, Y., & OBORNE, D. J. (1987). Pictorial communication of abstract verbs in relation to human computer interaction. *British Journal of Psychology*, **78**, 99-112.
- SANFELIU, M. C., & FERNANDEZ, A. (1996). A set of 254 Snodgrass-Vanderwart pictures standardized for Spanish: Norms for name agreement, image agreement, familiarity, and visual complexity. *Behavior Research Methods, Instruments, & Computers*, **28**, 537-555.
- SCHWARTZ, N. H., & PHILLIPPE, A. E. (1991). Individual differences in the retention of maps. *Contemporary Educational Psychology*, **16**, 171-182.
- SNODGRASS, J. G., & VANDERWART, M. (1980). A standardized set of 260 pictures. Norms for name agreement, image agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory*, **6**, 174-215.
- STANDING, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, **25**, 207-222.
- SUTTER, A., BECK, J., & GRAHAM, N. (1989). Contrast and spatial variables in texture segregation: Testing a simple spatial-frequency channels model. *Perception & Psychophysics*, **46**, 312-332.
- TREISMAN, A. (1986, November). Features and objects in visual processing. *Scientific American*, **255**, 114B-125.
- TREISMAN, A., & GELADE, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, **12**, 97-136.
- TREISMAN, A., & SOUTHER, J. (1985). Search asymmetry: A diagnostic preattentive processing of separable features. *Journal of Experimental Psychology: General*, **114**, 285-310.
- TULLIS, T. S. (1981). An evaluation of alphanumeric, graphic, and color information displays. *Human Factors*, **23**, 541-550.
- VASSILEV, A., & MITOV, D. (1976). Perceptual time and spatial frequency. *Vision Research*, **16**, 89-92.
- WALKER, R. E., NICOLAY, R. C., & STEARNS, C. R. (1965). Comparative accuracy of recognizing American and international road signs. *Journal of Applied Psychology*, **49**, 322-325.
- WANDMACHER, J., & AREND, U. (1985). Superiority of global features in classification and matching. *Psychological Research*, **47**, 143-157.