

How Many Pixels Do We Need to See Things?

Yang Cai

Human-Computer Interaction Institute,
School of Computer Science,
Carnegie Mellon University,
5000 Forbes Avenue, Pittsburgh, PA 15213, USA
ycai@cmu.edu

Abstract. Today's computer display devices normally provide more information than we need. In this paper, the author presents an empirical model that shows minimal pixel requirements for computer users to recognize things from digital photos under different contextual conditions. It is found that face recognition alone needs far fewer pixels than people normally thought. However, more pixels are needed for users to recognize objects within outdoor scenes and paintings. Color and age have effect on object recognition but the differences are not significant. The results can be applied to adaptive display design, computer vision, adaptive human-computer interaction and telecommunication system design.

1 Introduction

One of the challenges in today's human-computer interaction design is that the electronic components become smaller and smaller but users want the display to be larger and larger. Increasing the size of images will increase the data communication traffic and vice versa, since the bandwidth is normally limited. Also, processing large images would increase the computing time in orders of magnitude in those systems, such as machine vision, visualization, game engine, etc.

Studies [1] show that photos normally provide more information than we need. The redundancy can be as high as 99%! It is found out that the number of stimuli per display had no effect when the display time required to reach 75% accuracy was determined.[2] In facial communication, dramatic reductions in spatial and resolution of images can be tolerated by viewers.[3]

Therefore, from an economics point of view, the resolution in image transmission can be greatly reduced. For example, photos in newspapers normally only have two values for each dot (pixel): with ink or without ink. With grid screen processing, the size of the smallest pixels is increased so that the dots per area can be reduced greatly. However, the picture is still recognizable. Increasing the resolution of the grid screen can make the image more attractive but it doesn't increase the information content.

The resolution of an image can be represented by pixels. For example, how well subjective impressions of amount of architectural details can be predicated by objective measurement of the percentage of pixels covered by small elements.[4]

The famous face in Fig. 1 can be recognized in both resolutions a) 300 x 200 pixels and b) 150 x 100 pixels, but hardly recognized in c) 75 x 50 pixels. If we are asked to identify who is in the picture, then the 150 x 100 pixel image should be good enough. The background is redundant to the face recognition but certainly helpful in this case. Recent studies [5] show that lower resolutions of images actually are better for computer vision! For many high resolution images the process of finding the symmetry or the reflection plane of an object does not converge to the correct solution, e.g., the process converged to local minima due to the sensitivity of the symmetry value to noise and digitization errors. To overcome this problem, a multi-resolution scheme is often introduced, where an initial estimation of the symmetry transform is obtained at low resolution and is fine-tuned using high-resolution images.



a) image with resolution 300 x 200 pixels



b) 150 x 100 pixels



c) 75 x 50 pixels

Fig. 1. Examples of the redundancy in a Black and White image

For decades, computer scientists and engineers have been focused on high-end imaging and high-resolution display technologies. It seems that very few people are inter-

ested in the low-end imaging and low-resolution display technologies. In practice, scientists and engineers use *ad hoc* methods to come up the minimal pixel requirements. For example, 32x32, as a starting point.

From the visual cognition and vision study point of view, the question like “how many pixels do we really need to see things?” actually is an important one that is related to human pattern recognition, attention, and visual information processing. As we know, human visual attention is sensitive to the purposes of seeing and the demand of visual information resources is different from task to task. [6] Many pattern recognition processes are measured by reaction time or error rate. In this study, we use *number of pixels* as a measurement. Pixel is the smallest unit of an image element. In this study, we assume that all the pixels are square. The main goal of this study is to explore the limitations of minimal resolution for people to see things under various contextual conditions.

2 User Modeling Tool Design

The purpose of the tool for the lab experiment is to show: 1) the average minimal pixels of images (face, indoors, outdoors, etc.) that subjects can recognize things, 2) the effect of questions for subjects to determine the minimal pixels for face recognition, 3) the effect of age, and 4) the differences between the recognition with color images and black and white images.



Fig. 2. A screen shot of the experimental panel. (The size of the image on the cellular phone panel can be modified by pushing the buttons on the windows.)

The subjects were 19 university students and 4 faculty members. All subjects have had vision check and had no vision problems, such as color blindness, or low vision, etc. Ten unique images in both color and black & white formats were randomly chosen to cover 4 categories: (1) faces, (2) indoor scenes, (3) outdoors scenes, and (4) complex images. These images were also randomly ordered and presented individually with an interval via a simple computer program. For a color image, there are 24-bit colors for each pixel. For a gray image, there are 8-bit gray levels.

A Java-based software has been developed for a regular PC. The resolution of a hand-held prototype screen can be modified by pushing “Zoom In” or “Zoom Out” buttons. It allows a subject to enlarge a given image in the miniature display area *until* the threshold of a *correct* recognition is reached. The subject would then be asked to answer a question accompanying the given image. We did not include the false recognition data.

The program consisted of a main image area that displayed the given image initially at 10x10 pixels. Subjects were asked to “zoom in” on the image using a button on the bar at the bottom of the window until the moment he/she recognized the contents of the image. The subjects were also asked to answer and type in their replies to a question presented in the upper left corner immediately upon image recognition. The subjects were allowed to continue zooming in on the image until the question could be answered. The subjects proceeded through the 20 images (labeled 0 through 19) by use of the “Next” button in the upper right corner of the window.

We assumed that given a set of randomly selected images, those containing human faces would be recognized at smaller resolutions, followed by simple commonly known objects, and then more-complex indoor and outdoor scenes. Regarding facial recognition, we believed that simple recognition of a face would require the simplest features, while gender identification and recognition of a well-known individual (i.e. former President Bill Clinton) would require more pixels. We also assumed that the subject’s age had no effect on required image size, and that an image’s being in black & white or color would make a negligible difference, with a slight advantage toward color images.

3 Results

First, we asked subjects “What is this?” about photos of face-only, indoors, outdoors, figures, and complex scenes (such as oil paintings). Subjects adjusted the size of the images until they could recognize in the image. Facial recognition required significantly fewer pixels than for human figures, indoor scenes, and outdoor scenes. As expected, complicated scenes required the largest number of pixels for identification. See the data in Table 1 and Fig. 3.

Second, we tested a set photos of faces and asked subjects with three questions: “Who’s this?” “What is this?” “Male or female?” respectively. The results show that the minimal resolution in corresponding to the question “Who’s this?” is the smallest. To identify male or female needs more resolution, since it’s hard to distinguish in

many cases in real world. To some extent, the number of pixels reflects the difficulty of the cognitive task. See Table 2 and Fig. 4 for the results.

Table 1. Minimal Pixels for Identifying Objects

Catalog	Minimal Size	Minimal Pixels
Face	17 x 17	289
Outdoor	32 x 32	1024
Figure	35 x 35	1225
Indoor	40 x 40	1600
Complex	47 x 47	2209

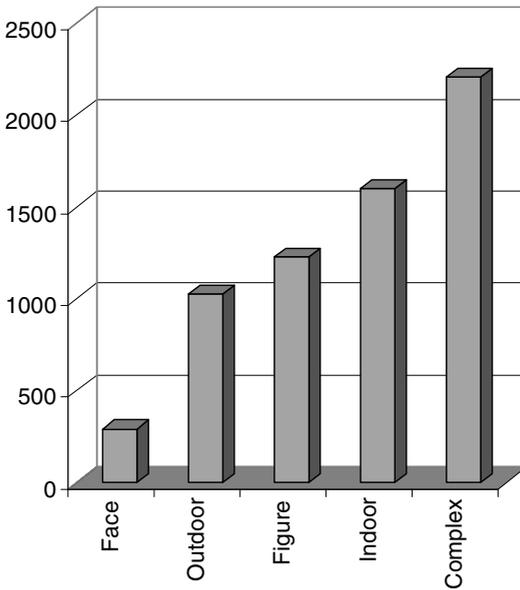


Fig. 3. Minimal pixels for object recognition

Third, we showed subjects images with color or black and white to compare the differences. Given that black & white and their color counterparts were randomized in order of presentation, the subject’s short-term memory would not have altered these findings. Interestingly enough, black and white images need fewer pixels than color. See Table 3 and Fig. 5 for the test results.

Finally, we showed the images to different age groups to see whether age is an effect for determining the minimal resolution of images. We use age 21 as a cutting point, since it is a normal line to separate undergraduate students and post-graduate

students as well as other “adults”. It is amazing to find that younger subjects actually use more pixels to recognize objects than elder subjects. Experience and patience might play a role here. However, the differences are not significant statistically. See Table 4 and Fig. 6 for the test results.

Table 2. Minimal Pixels for Identifying Faces

Question	Minimal Size	Minimal Pixels
Who is this?	17 x 17	289
What is this?	32 x 32	1024
Male or female?	35 x 35	1225

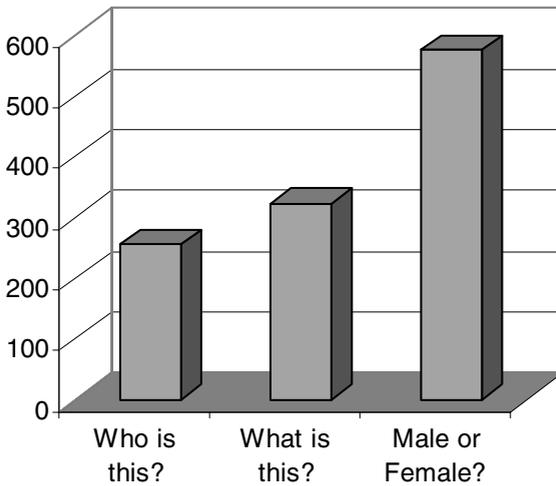


Fig. 4. Minimal pixels for facial recognition under various inquiries

4 Conclusions

In this paper, the author presents an empirical study of the minimal resolution of images in terms of pixels for computer users to recognize visual objects. Here are preliminary conclusions:

Table 3. Color images versus B&W images

Image No. Color / B &W	Color Image Minimal Size	B & W Image Minimal Pixels
[0] / [19]	40 x 40	30 x 30
[8] / [1]	20 x 20	18 x 18
[2] / [14]	63 x 63	38 x 38
[3] / [17]	31 x 31	21 x 21
[4] / [6]	42 x 42	20 x 20
[5] / [10]	35 x 35	31 x 31
[7] / [13]	36 x 36	21 x 21
[9] / [18]	36 x 36	19 x 19
[12] / [16]	15 x 15	12 x 12
[15] / [11]	28 x 28	46 x 46

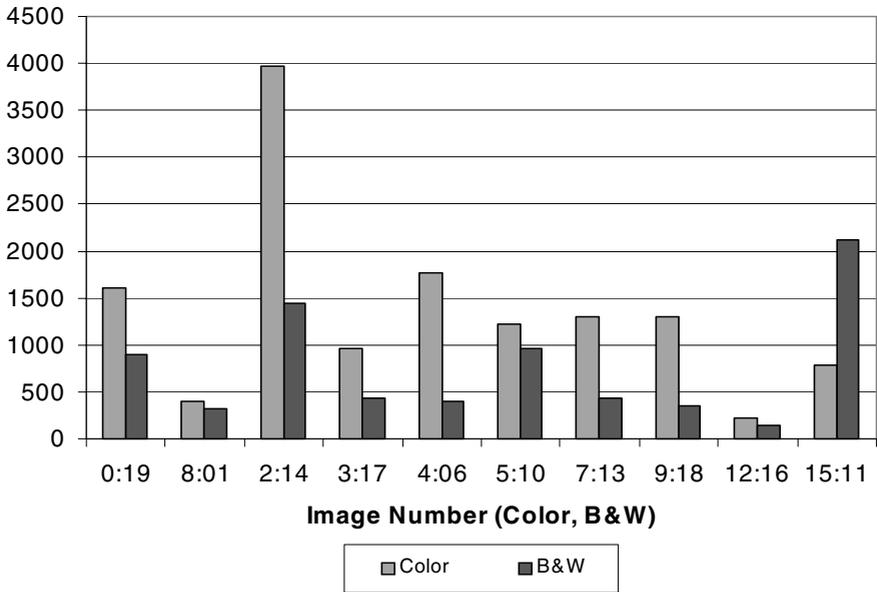


Fig. 5. Minimal pixels for color vs. B&W

Table 4. Minimal sizes versus ages

Image	Less than age 21	Age 21 or elder
[0]	42 x 42	38 x 38
[2]	75 x 75	49 x 49
[3]	32 x 32	31 x 31
[4]	15 x 15	14 x 14
[5]	40 x 40	30 x 30
[7]	25 x 25	21 x 21
[8]	21 x 21	19 x 19
[9]	40 x 40	31 x 31
[12]	17 x 17	14 x 14
[15]	29 x 29	28 x 28

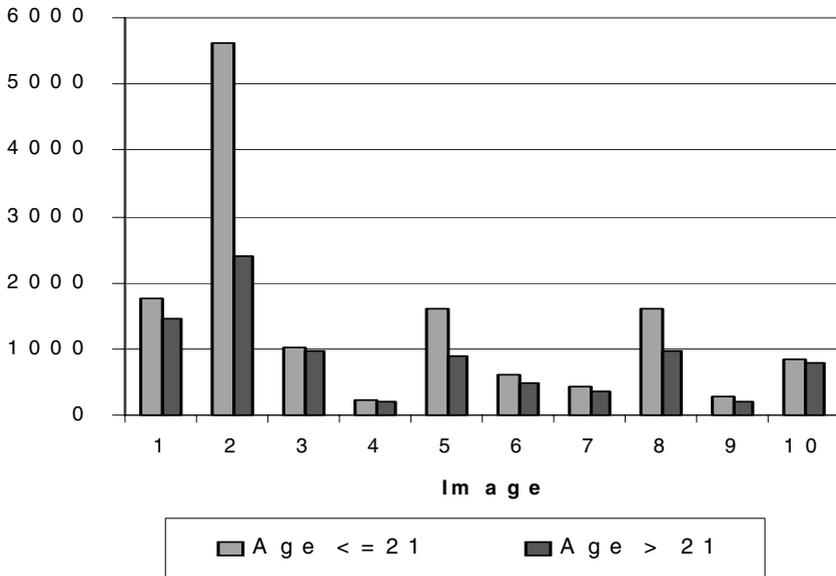


Fig. 6. Minimal pixels for age groups

First, face recognition needs far fewer pixels than people normally thought, especially, if the subject knows what he/she is looking for. The experiment agrees with the theory that human visual attention is sensitive to the purposes of seeing and the demand of visual information resources is different from task to task. [6] When we think about our faces, they are well-structured compared to other objects. They are also highly featured so that it is easy to identify a face from an image relatively. Also, it's still under investigation, whether or not a human has special "wired connections" to recognize a face.

Second, although it is context dependent, we found a general trend of the order of the complexity of human visual information processing. The order, from less pixels to more pixels, is "faces", "outdoors", "figure", "indoors", and "complex scenes." Complex scenes, such as oil paintings contain more vague objects that confuse viewers and make it hard to identify things.

Third, we found that "pixel" can be a numerical measurement of visual information processing. Traditionally, cognition scientists use reaction time, number of entities, error rate, etc. to measure the visual information processing. Pixel is a simple way to capture and compute within normal human-computer interaction environment. However, pixels of an image may not always represent the amount of visual information, because there are redundant pixels in a picture if we don't preprocess the image carefully. For example, for face recognition tasks, we cut off the background that is outside the face outline. Also we used a square image to simplify the measurement.

Fourth, subjects need slightly fewer pixels to recognize things with black and white images than color images. However, those differences are not statistically significant.

Fifth, mature subjects (age 21 and up) appeared to need less pixels than younger ones in recognition tasks. However, we need more data to prove this finding. One explanation is that mature subjects have more visual experience and more stability in perception, given the same eyesight.

The purpose of this study is to move toward the adaptive display design with adjustable picture resolutions. There is a wide range of possibilities to apply the heuristics from this study, for example, the adaptive display on mobile phones. As we know, the bandwidth of wireless data transmission is very limited. If the chip in a mobile phone knows the context of the transmitted video, it might be able to change the resolution adaptively so that it can save the bandwidth dramatically. In addition, the results can be a direct reference for computer vision study because it shows how few pixels a human subject needs to recognize things in a picture. In the future, computer vision systems might not need a high-end image acquisition and processing system, if it's designed to solve a specific problem.

Acknowledgement. The author thanks to Mr. Peter Hu for his assistance in data collection and data processing.

References

1. Strobel, L, et al, *Visual Concepts for Photographers*, ISBN 7-80007-236-3,1997
2. Petersik, J.T. *The Detection of Stimuli Rotating in Depth Amid Linear Motion and Rotation Distractors*, *Vision Research*, August 1996, vol.36, no.15, pp.2271–2281(11)
3. Bruce, V. *The Role of the face in communication: implications for video phone design*, *Interaction with Computers*, June 1996, vol. 8, no.2, pp. 166–176(11)
4. Stamps III, A.E. *Architectural Detail, Van der Laan Septaves and Pixel Counts*, *Design Studies*, January 1999, vol.20, no.1, pp. 83–97 (15)
5. Zabrodsky, H. and et al, *Symmetry as a Continuous Feature*, *IEEE Trans. On Pattern Analysis and Machine Intelligence*, Vol. 17, No.12, December, 1995
6. Solso, R, *Cognition and Visual Art*, The MIT Press, 1999
7. Brand, S. *The Media Lab*, Penguin Books, 1988, pp. 170–172
8. Cai, Y., *Pictorial Thinking*, *Cognitive Science*, Vol.1, 1986
9. Cai, Y., *Texture Measurement of Visual Appearance of Recognition of Oil Paintings*, *IEEE IMTC*, Anchorage, May, 2002
10. Buhmann, J.M., et al, *Dithered Color Quantization*, *Computer Graphics Forum*, August 1998, vol. 17, no.3, pp.219–231
11. Batchelor, D. *Minimalism – Movements in Modern Art*, University of Cambridge, 1997

Appendix A. Test Images



Image 0 Color



Image 1 B&W



Image 2 Color



Image 3 color



Image 4 color



Image 5 color



Image 6 B&W



Image 7 Color



Image 8 Color



Image 9 Color



Image 10 B&W



Image 11 Color



Image 12 Color



Image 13 B&W



Image 14 B&W



Image 15 Color



Image 16 B&W



Image 17 B&W



Image 18 B&W



Image 19 B&W