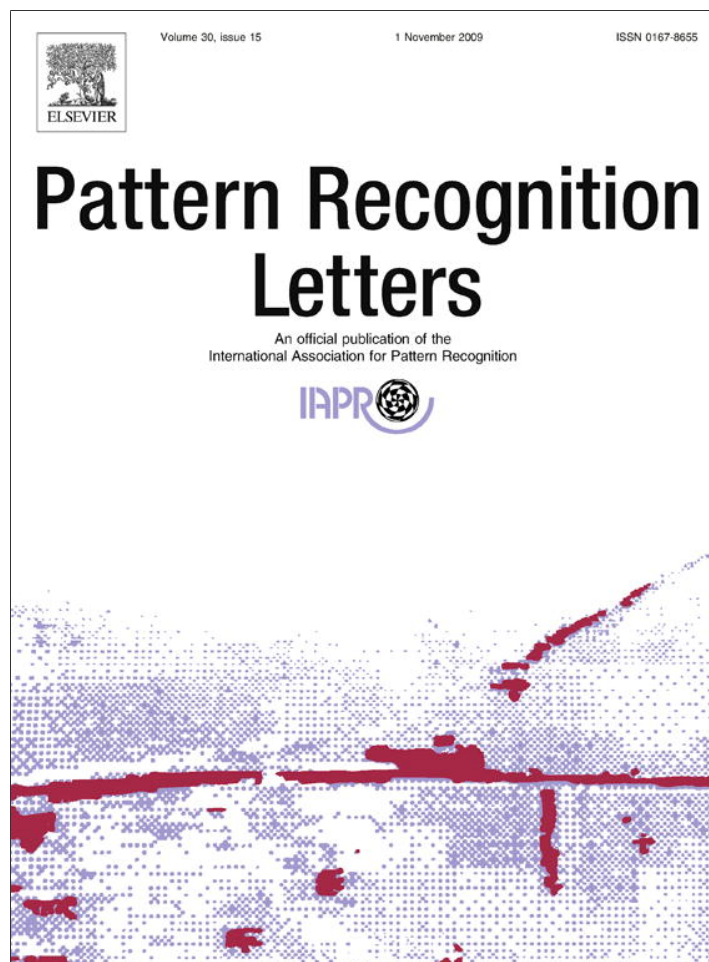


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

The number of all possible meaningful or discernible pictures

Theo Pavlidis*

Department of Computer Science, Stony Brook University, Stony Brook, NY, 11794, United States

ARTICLE INFO

Article history:

Received 7 April 2009

Received in revised form 29 May 2009

Available online 12 July 2009

Communicated by T.K. Ho

Keywords:

Statistical pattern recognition

Image retrieval

ABSTRACT

The paper provides lower bounds on the number of all possible images that appear meaningful to a human observer as well as the number of all possible images that appear pair-wise distinct to a human observer. These numbers suggest that it is impractical to construct training or testing sets of images that are dense in the set of all images unless the class of images is restricted.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

One of the challenges in statistical pattern recognition is to select representative sets of images to train and test an algorithm. A similar challenge occurs in content-based image retrieval where one must develop a measure of image similarity. Again there is a need for representative image sets for training and learning. Quite often authors claim (explicitly or implicitly) that the set they use is representative because of the large number of images it contains, usually tens of thousands or even several million of images.

The purpose of this paper is to show that using even millions of images is no guarantee for having a representative set because the number of all possible images is in the trillions. Unless one can specify characteristics of the set of images of interest (for example, images of numerical digits), sheer numbers do not guarantee that a set is representative of all images.

At first sight, the enormity of the number of all possible images seems obvious. For example, consider 32 by 32 images with 3 bits per pixel (one bit per color). The number of possible images of such size is 2^{3072} or about 10^{300} . The number of images with three bytes per pixel and 256 by 256 size is about $10^{500,000}$. The trouble with such arguments is that most of these images do not represent scenes that could possibly have been captured by a camera. Therefore we must ask a much harder question.

2. The number of meaningful (valid) pictures

What is the number of all possible pictures that may represent a scene? In other words, pictures that will represent not just a collec-

tion of random pixels but a collection of pixels that, at least, some human observers will interpret as representing something that could have been captured by a camera. I used for brevity the term *meaningful pictures* even though the word “meaning” depends a lot on the observer. (Others may prefer the term *valid pictures*.)

I will try to answer this question by providing a lower bound. Suppose I subdivide an image in K -by- K square blocks and paint each block with one of a small set of specific images. Fig. 1 shows an example where K equals 9 (for a full color and full size version of the figure see <http://www.theopavlidis.com/technology/Image-Number/index.htm>.) Two images are shown side by side and they differ in only one block. Each block may be either empty or contain an image of a sea creature. The block size is 60 by 60, so the size of each image is 540 by 540 and can be displayed both on a page side by side and this is the only reason for selecting K equal to 9.

Each image is meaningful (valid) to a human observer (although people may have trouble finding the block where the two images differ). How many such images exist? Because each block can have one of five configurations and there are 81 such blocks the answer is $5^{81} > 10^{56}$. (A trillion is 10^{12} .) The number 10^{56} is only a lower bound on the number of all possible valid images. By selecting K equal to 20 and the same block size we would have a 1200 by 1200 pixel image that fits easily on a computer screen. We could easily accommodate 10 choices for each block, so the number of such images would be 10^{400} and we could go even higher. And that would still leave many images out.

Estimate No. 1: The number of meaningful/valid images on a 1200 by 1200 display is at least as high as 10^{400} . The number of images corresponding to the display of Fig. 1 is 10^{56} , a very conservative lower bound to the number of all possible meaningful/valid images.

* Fax: +1 516 632 8334.

E-mail address: t.pavlidis@ieee.org

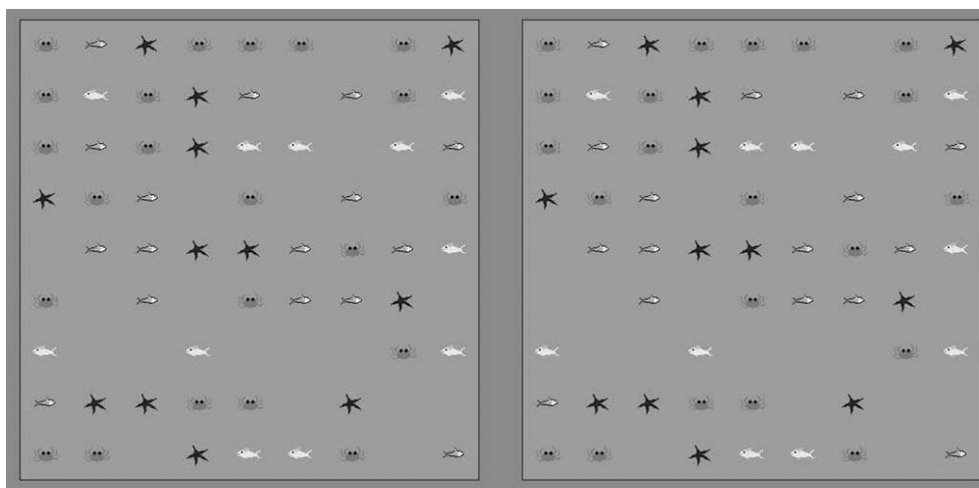


Fig. 1. Sample of two 9-by-9 arrangement of blocks that may contain a picture of one of four sea-creatures or be empty. Both images are valid and meaningful to a human observer, although it may be hard to differentiate between the two unless one peruses the images carefully.

3. The number of discernible pictures

An even harder question is to ask how such images will be *discernible* by humans. For example, there may be several images of foliage but most people could not tell them apart. This question can be further refined by distinguishing between pairs of pictures that are seen as *different when displayed side by side* or pairs that can be *differentiated from memory*. In this note I will consider only the first case. Clearly, the number of images that can be distinguished from memory is much smaller. For example, there are about five billion people in the world and it is likely that pictures of any two individuals, when placed side by side could be differentiated. But if a person is shown two pictures separately with, say, an hour gap between displays then the number is going to be probably close to 1000 rather than the billions that are discernible when displayed side by side. I wrote a program that uses a random number generator to create K -by- K boards and then (randomly) changes the images of one of the squares and displays both boards side-by-side. An executable module as well as the source code (as a zipped Microsoft Visual Studio project) are available at <http://www.theopavlidis.com/technology/ImageNumber/index.htm> (Fig. 1 was produced by this program).

The code of the program is quite simple. If M is the number of possible choices for each block, then one just needs to fill a K -by- K array with random numbers in the $[0, M - 1]$ interval. To create the second image one selects at random an element of the array and selects at random from of the $M - 1$ values, other than the current value of the element.

The complete program allows for timing of the user reaction and readers are invited to experiment by going to the site whose link was given above and downloading the program. The program keeps track of the average time for their choices. For K equal to 6 the time is in the order of a few seconds, from about 4 s for two configurations to about 6 s for five configurations. (I have asked several people to do the task from printouts of images and they all did it within a few seconds.) Readers can confirm these numbers by experimenting themselves.

Thus 5^{36} seems to be a lower bound on the number of discernible images. That exceeds 10^{25} , which is greater than a trillion squared. If we allow observers to spend over 10 s perusing a pair of images then they find differences for higher values of K . (Even for K equal to 9, as in Fig. 1, it takes less than 30 s.)

Estimate No. 2: 10^{25} (greater than a trillion squared) is a very conservative lower bound to the number of all possible discernible images.

The ability of humans to differentiate images on the basis of small details, even when recalled from memory, has been documented recently by Brady et al. (2008), therefore Estimate No. 2 of lower bounds is not that surprising.

4. Conclusions

The huge number of possible images raises several questions about pixel based vision both from computational and neuroscience viewpoints. An editorial in *Nature* by Mazer (2008) discusses the issues from the viewpoint of neuroscience pointing out existence of areas in the brain that respond to structures rather than pixels. (Additional references can be found in that editorial.)

Some researchers in machine vision dismiss the issue by arguing that humans lump large numbers of images into a small number of categories, but that misses an important point. Algorithms for image classification or retrieval perform their computations on the pixel space and one must prove that all these calculations provide similar results on widely differing pixels sets that correspond to the same category. The wide disparity in pixels values of “similar” images (and conversely pixel-wide similarities on “different” images) present obstacles in extending laboratory results to practice (Pavlidis, 2008).

Other computer vision authors have focused on the challenges posed by the ability of the human visual system to process very large amounts of information and its implication on learning mechanisms. It is beyond the scope of this short paper to provide a review of all such efforts but I want to point out one of the earlier papers by Tsotsos (1987). While an explicit number of all possible images is not the focus of that paper, the parameters discussed in it point to an estimate of 10^{5400} , much higher than Estimate No. 1.

Restricting the classes of images to be dealt with simplifies, but it does not eliminate the problem. The main benefit from class restriction is the possibility to model the formation of images and therefore use synthetic data. This has been done successfully in the area of document analysis where some researchers have been relying on synthetic data to augment the size of the data sets used to train and test classifiers (Baird, 2006). In a recent paper Nonnemaker and Baird (2009) have used the typeface description model of Metafont and an image degradation model to show that properly constructed synthetic data are not only safe to use but also improve performance. The paper by Jin and Geman (2006) presents a novel approach in dealing with the high dimensionality of the image space by relying on context at several levels.

From a practical viewpoint, the synthetic data approach is the simplest to implement in order to deal with the issue raised in this paper. For example, when dealing with image retrieval problems the “learning” and “testing” sets should be augmented by images obtained from the given by simulating different camera settings and lighting conditions. Such transformations are simple to implement and can increase the size of the sets by at least a factor of 10. On the other hand, instead of creating synthetic images one might use features such as FIST (Lowe, 2004) that are invariant under several transformations.

Clearly, more systematic research could establish tighter bounds than those given here, but such results may be more important for psychology or neuroscience than technology. Whether the number of all possible images is 10^{30} or 10^{3000} may be of interest to pure science, but either number implies that it is impossible to conduct research that relies on “knowing” all possible images.

Acknowledgements

I want to thank Prof. G. Zelinsky (Dept. of Psychology, Stony Brook University) for helpful comments on an earlier draft and for pointing out Brady et al. (2008).

References

- Baird, H.S., 2006. The state of the art of document degradation modeling. In: Chaudhuri, B.B. (Ed.), *Digital document processing: Major directions and recent advances*, Advances in Pattern Recognition Series. Springer, pp. 261–279.
- Brady, T.F., Konkle, T., Alvarez, G.A., Oliva, A., 2008. Visual long-term memory has a massive storage capacity for object details. *PNAS* 105 (38), 14325–14329.
- Jin, Y., Geman, S., 2006. Context and hierarchy in a probabilistic image model. *CVPR*, 2145–2152.
- Lowe, D., 2004. Distinctive image features from scale-invariant keypoints. *Internat. J. Comp. Vis.*, 91–110.
- Mazer, J.A., 2008. So many pixels, so little time. *Nat. Neurosci.* 11 (November), 1243–1244.
- Nonnemaker, J., Baird, H.S., 2009. Using synthetic data safely in classification. In: *Proc. IS&T/SPIE Conf. on Document Recognition and Retrieval (DRR 2009)*, San Jose, CA, January 28–February 1. Paper 72470G.
- Pavlidis, T., 2008. Limitations of content-based image retrieval. In: *Invited Plenary Talk at the 19th Internat. Conf. on Pattern Recognition*, Tampa, Florida, December 8–11. Slides can be found in <http://www.theopavlidis.com/technology/CBIR/PaperB/icpr08.htm>.
- Tsotsos, J.K., 1987. A ‘complexity level’ analysis of vision. *ICCV*, 346–355.