WAYS OF MACHINE SEEING[1]

GeoffCox

You are looking at the front cover of the book Ways of Seeing written by John Berger in 1972. The text is the script of the TV series, and if you've seen the programmes, you can almost hear the distinctive pedagogic tone of Berger's voice as you read his words: "The relation between what we see and what we know is never settled."[2]

The image by Magritte on the cover further emphasises the point about the deep ambiguity of images and the always-present difficulty of legibility between words and seeing.[3] In addition to the explicit reference to the "artwork" essay by Walter Benjamin,[4] the TV programme employed Brechtian techniques, such as revealing the technical apparatus of the studio; to encourage viewers not to simply watch (or read) in an easy way but rather to be forced into an analysis of elements of "separation" that would lead to a "return from alienation".[5] Berger further reminded the viewer of the specifics of the technical reproduction in use and its ideological force in a similar manner: "But remember that I am controlling and using for my own purposes the means of reproduction needed for these programmes [...] with this programme as with all programmes, you receive images and meanings which are arranged. I hope you will consider what I arrange but please remain skeptical of it."

That you are not really looking at the book as such but a scanned image of a book — viewable by means of an embedded link to a server where the image is stored — testifies to the ways in which what, and how, we see and know is further unsettled through complex assemblages of elements. The increasing use of relational machines such as search engines is a good example of the ways in which knowledge is filtered at the expense of the more specific detail on how it was produced. Knowledge is now produced in relation to planetary computational infrastructures in which other agents such as algorithms generalise massive amounts of (big) data.[6]

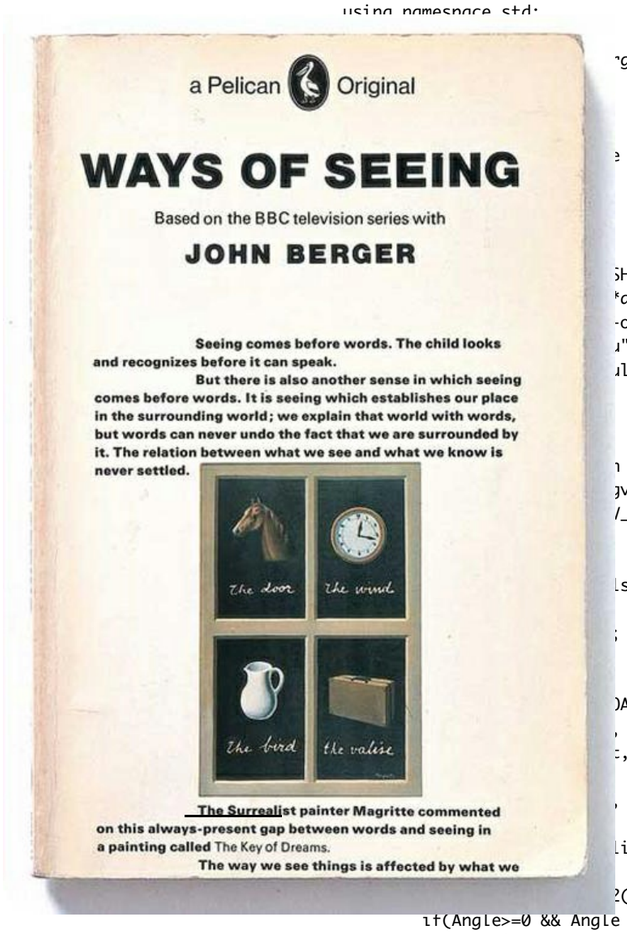
Clearly algorithms do not act alone or with magical (totalising) power but exist as part of larger infrastructures and ideologies. Some well-publicised recent cases have come to public attention that exemplify a contemporary politics (and crisis) of representation in this way, such as the Google search results for "three black teenagers" and "three white teenagers" (mug shots and happy teens at play, respectively).[7]

There is a sense in which the world begins to be reproduced through computational models and algorithmic logic, changing what and how we see, think and even behave. Subjects are produced in relation to what algorithms understand about our intentions, gestures, behaviours, opinions, or desires, through aggregating massive amounts of data (data mining) and machine learning (the predictive practices of data mining).[8] That machines learn is accounted for through a combination of calculative practices that help to approximate what will likely happen through the use of different algorithms and models. The difficulty lies in to what extent these generalisations are accurate, or to what degree the predictive model is valid, or "able to generalise" sufficiently well. Hence the "learners" (machine learning algorithms), although working at the level of generalisation, are also highly contextual and specific to the fields in which they operate in a coming together of what Adrian Mackenzie calls a "play of truth and falsehood".[9]

Thus what constitutes knowledge can be seen to be controlled and arranged in new ways that invoke Berger's earlier call for skepticism. Antoinette Rouvroy is similarly concerned that algorithms begin to define what counts for knowledge as a further case of subjectivation, as we are unable to substantively intervene in these processes of how knowledge is produced.[10] Her claim is that knowledge is delivered "without truth" through the increasing use of machines that filter it through the use of search engines that have no interest in content as such or detail on how knowledge is generated. Instead they privilege real-time relational infrastructures that subsume the knowledge of workers and machines into generalised assemblages as techniques of "algorithmic governmentality".[11]

In this sense, the knowledge produced is bound together with systems of power that are more and more visual and hence ambiguous in character. And clearly computers further complicate the field of visuality, and ways of seeing, especially in relation to the interplay of knowledge and power. Aside from the totalizing aspects (that I have outlined thus far), there are also significant "points of slippage or instability" of epistemic authority,[12] or what Berger would no doubt identify as the further unsettling of the relations between seeing and knowing. So, if algorithms can be understood as seeing, in what sense, and under what conditions? Algorithms are ideological only inasmuch as they are part of larger infrastructures and assemblages.

But to ask whether machines can see or not is the wrong question to ask, rather we should discuss how machines have changed the nature of seeing and hence our knowledge of the world.[13] In this we should not try to oppose machine and human seeing but take them to be more thoroughly entangled — a more "posthuman" or "new materialist" position that challenges the onto-epistemological character of seeing — and produces new kinds of knowledge-power that both challenges as well as extends the anthropomorphism of vision and its attachment to dominant forms of rationality. Clearly there are other (nonhuman) perspectives that also illuminate our understanding of the world. This pedagogic (and political) impulse is perfectly in keeping with Ways of Seeing and its project of visual literacy.[14]




```
// tess.cpp:
// Recognize text on an image using Tesseract API and print it to the screen
// Usage: ./tess image.png

#include
#include
#include
#include
#include
#include "opencv2/core/core.hpp"
#include "opencv2/features2d/features2d.hpp"
#include "opencv2/highgui/highgui.hpp"
#include "opencv2/imgproc/imgproc.hpp"
#include

using namespace cv;

using namespace std;

                                                    gv)


                                                    e specify the input image!" << std::endl;


                                                    SHEY_SIMPLEX;
                                                    *api = new tesseract::TessBaseAPI();
                                                    -ocr with English, without specifying tessdata pa
                                                    ")) {
                                                    uld not initialize tesseract.\n");


                                                    n leptonica library
                                                    gv[1]);
                                                    /_LOAD_IMAGE_COLOR);


                                                    ls, im.type(), cv::Scalar(0));


                                                    OAD_IMAGE_GRAYSCALE);
                                                    , 3 );
                                                    :, CV_GRAY2BGR );


                                                    , 1, CV_PI/180, 80, 80, 10 );


                                                    lines.size(); i++ )

                                                    ?(lines[i][3]- lines[i][1], lines[i][2]- lines[i]
    if(Angle>=0 && Angle <=5){
        numlines++;
        line( color_dst, Point(lines[i][0], lines[i][1]),
              Point(lines[i][2], lines[i][3]), Scalar(0,0,255), 3, 8 );
    }
}
if(numlines>40){
    //Mat screen=cvCreateMat(im.rows,im.cols,im.type());
    api->SetImage(image);
    api->Recognize(0);
    //outText = api->GetUTF8Text();
    tesseract::ResultIterator* ri = api->GetIterator();
    tesseract::PageIteratorLevel level = tesseract::RIL_SYMBOL;

    if (ri != 0) {
        do {
            const char* word = ri->GetUTF8Text(level);
            float conf = ri->Confidence(level);
            int x1, y1, x2, y2;

            ri->BoundingBox(level, &x1, &y1, &x2, &y2);
//              printf("word: '%s';  \tconf: %.2f; BoundingBox: %d,%d,%d,%
//                     word, conf, x1, y1, x2, y2);
            bool ignore;
            int psize,fid;
```

The problem is one of learning in its widest sense, and “machine learning” techniques are employed on data to produce forms of knowledge that are inextricably bound to hegemonic systems of power and prejudice.

What is required is an expansion of this ethic to algorithmic literacy to examine how machine vision unsettles the relations between what we see and what we know in new ways.