Guide to Reading Lowe's SIFT paper

Leo Dorst

May 4, 2017

1 Appendix B.1.1 Repeated

Whenever you have a scientific paper that is not quite in your field (yet), it is useful to read Abstract, Introduction, and Conclusions first. Together, these should provide a roadmap for the paper, and help prevent you getting lost in it.

Then, read the paper through once, trying to understand it, but dont spend too much effort on things you dont get immediately. Often these will become clear later on. There may be various reasons for this:

- The paper may be badly written, using things before they are properly introduced, or using inconvenient symbols or concepts. Also, the main line and the side tracks are not always clearly separated.
- The paper may use strange abstractions for very concrete things, without you (or the author) actually needing them at the abstract level. This is sometimes done to impress.
- There may be excursions that are interesting to specialists, but not required for understanding the paper at the level you need it.
- The author may assume common knowledge you dont have (yet).
- The paper may be the wrong paper to read for what you want to know; but when you see the author refer to other material in context you may be able to home in on the right information anyway.

The first pass of reading will help you understand the structure and identify the various difficulties you will have in understanding the paper, sometimes even already resolving them.

When you have decided that this is really the paper you want to read, and where the interesting bits are for you, you read it again. In this second pass, you should really make an effort to understand everything relevant (you will have identified the relevant parts in pass one). Follow the reasoning in detail, fill in the blanks using the references, some background material (including internet search for mathematical terms you may not know). This is not a linear pass, you have to deconstruct the paper! This pass should be very rewarding, you are increasing your knowledge considerably in doing this. (Always scout the references for useful literature you may not yet know!) Make notes in the margin, or attach your notes to the paper. You dont want this effort to get lost, so that you might have to do it again... At the end of this pass, you should know the relevant things well enough to be able to explain them to others.

After you have read and understood it and have thought about it for a few days or weeks, go back a third time. You will have the satisfaction of seeing it all fall into place, and often some of the obscure or more abstract bits are now suddenly clear, deepening your understanding of the subject and the field. Summarize the salient points in 5 or so bullets on the title page, so that when you see the paper again you will remember what you liked about it (or didnt).

In brief, personalize the paper so that it becomes part of your ever-expanding knowledge.

2 How Lowe Can You Get?

This paper is a classic, but it is not particularly well written. Implementation considerations are interwoven with theoretical explanations - we distinctly get the feeling he is an implementer. But it is precisely because he implemented this, and because it worked to do things we could not do before, that people use his ideas.

Abstract

The abstract should be clear enough - can you already relate this to your mosaicing problem?

1. Introduction

An introduction should contain motivation and an overview of the method. And it does! Understand what each paragraph is saying about method or use.

2. Related research

Research should progress, so it is important that the author shows the reader what was done before with more or less the same goals, and how (according to the author) this did not resolve the problem fully; or how earlier work contained some good ideas that the author will use. If you are in the field, you will recognize the references; if not, the problem the author actually addresses should become clearer as he uses them to define his approach. You do not need to look up all these references; but if you are in the field, you may find some unexpected things you did not know about, and which may be relevant to you.
Highlight a few (one or two) terms per paragraph as a means to make yourself aware of its purpose.

3. Detection of scale-space extrema

And we're off! Scale space is a bit new to you at this point, look it up or come to the lecture. But you know Gaussian convolution!
Your first pass of reading should tell you that the order of this section is messy. He want to use his own method (extrema of difference of Gaussians) because it is efficient. Oh, and actually it is a good approximation the Laplacian (look up what that is!). That is of course the really fundamental relationship to local structure. After having given (1), he derives it (he could/should have done that first), but does not quite return to 'his' D. A lab question asks you to close the gap.
What is an octave? Why does he 'need $s + 3$ images in the stack of blurred images for each octave'?

3.1 Detection of scale-space extrema

It is a pity that he does not quite give the intuition of why you want these extreme in 'all' directions. This must be standard scale space theory. Look it up. What does this mean for the kind of detection he does in the local structure, can he detect blobs, ridges? Remember, D is essentially the Laplacian.
But he is taking a very simple local way of determining an extremum. Which? Because this is so simple, he is concerned about how good it is, and that leads into 3.2 and 3.3.

3.2 Frequency of sampling in scale

We are diving into details here - he is doing some experiments, for what purpose? What is the outcome?

3.2 Frequency of sampling in the spatial domain

What is going on here? He is going to smooth before he detects extrema. But he has already smoothed to make the scale space images. Why is yet another smoothing needed? Could he not have combined that with the earlier smoothing?

4. Accurate Keypoint Localization

Finally, some math! Nothing could be clearer and unambiguous. Note how he used to do it naively but begins to appreciate that he is really in the realm of local structure, allowing sub-pixel accuracy.
Equation (2) is familiar, equation (3) you should be able to derive (in some years, it is even part of an exam!), and the consequence (4) you should derive for yourself.
He needs to do all this on the Laplacian or D image. And then, on page 11, he is doing the Hessian and derivative of D by local differences. Is that consistent? Do you realize that these are *third* and *fourth* derivatives of scale space image data?!

4. Accurate Keypoint Localization

He is going to do more detailed analysis of spatial peaks of D at a given scale. This is actually a slightly different way of looking at the curvature gauge: we have treated the eigenvectors and eigenvalues; he uses trace and determinant. Why?
Why is it 'unlikely' that the determinant is negative (I would not know), but why does that not bother him?
Count the number of floating point operations - do you get about 20? If not, what does he also include in his calculation?

5. Orientation assignment

The features are not going to orientation invariant, but the method will be. Resolve that paradox in your mind.
In the formula, he does not use the `atan2` function. I think he should (and probably does). Why?
He computes the gradients - where in his total set of computations is that done do you think - everywhere or only at the peaks? And is this still the gradient of D (as above) or of something else?
'Finally a parabola is fit', make sure you understand to what. This is in fact a 1D version of the localization of extrema we saw before.
Again he puts in some experiments for this step. We can feel him developing and testing his modules step by step. Educational it is.

6. The local image descriptor

We get a biological motivation. This seems a strange place for it, you could also imagine it in the introduction since we AI people would have experienced that as an extra reason to be interested in the method. It must have been an afterthought.

6.1 Descriptor representation

Here we get a descriptor, localized by a Gaussian of width σ. What is this σ, is it one of the earlier values given?
What is trilinear interpolation and why are we in 3D? What are the 3 dimensions?
What does he mean by 'affine changes in illumination'? Gather this from the surrounding text!
That bit about 0.2: yuck, how *ad hoc*!

6.2 Descriptor testing

We get the motivation for some numbers he mentioned earlier. This and the next sections are of course important to convince people of the effectiveness of his method, but they do not make for exciting reading.

6.3 Sensitivity to affine change

Yawn. I mean, impressive enough.

6.4 Matching to large databases

For modern applications, these properties of SIFT (and its successors) are very relevant.

7 Application to obect recognition

7.1 Keypoint matching

7.2 Efficient nearest neighbor matching

7.3 Clustering with the Hough transform

The Hough transform is also part of Beeldverwerken. He uses elementary techniques, but this papaer enabled things we could not do before, certainly not as robustly.

7.4 Solution for affine parameters

What is an orthographic projection?
This section has techniques that should look very familiar indeed!

8 Recognition examples

9 Conclusions

You read these in your first pass. Do you agree with his conclusions? Are there some that are missing?