

# Multimedia information retrieval: Homework assignment 3

*Machine used: Huisuil03*

Gideon Hanse, S1630784

March 12, 2019

## 1. Salient Point Detector: SIFT

### (a) Compare the notredame images

I used the getsift executable that was provided with the assignment to detect the salient points in two different pictures of the notredame, notredame1.jpg and notredame2.jpg. The main difference between the pictures is that notredame2.jpg is more of a close up than notredame1.jpg. Running the sift salient point detector with the notredame images provided different salient points for both pictures. Salient points in the pictures could be considered to be edges, windows, ornamentals. Although some similar salient points are found in both images, there are some differences. In general, the sift algorithm seemed to perform better in the closed up picture of the notredame as many of the very recognizable parts of the church are classified as salient points. This could probably be due to the noise of the passersby in the picture that was taken from farther away. Figure 1 contains a comparison between the salient points found in both images.

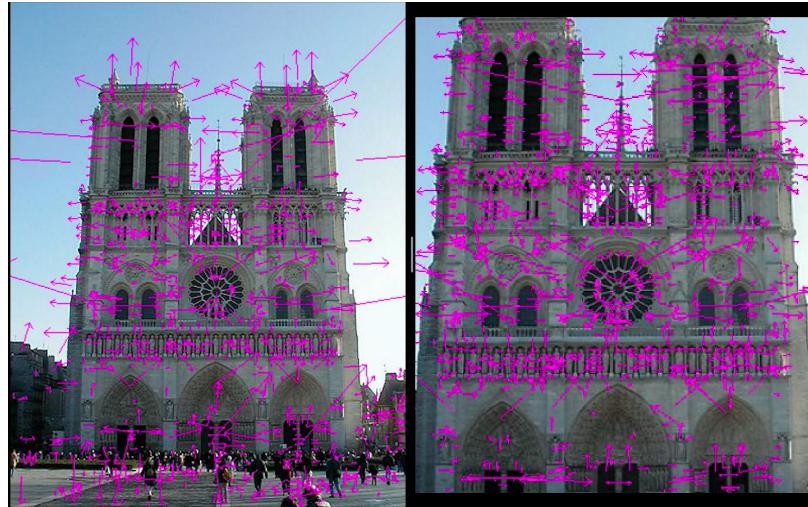


Figure 1: Salient points in both notredame images

### (b) Compare matching methods

The salient point detector seems to be quite accurate, as two pictures from Tom Cruise taken from a different angle are matched quite well. As presented in figure 2, both the simple and the advanced matching methods are successful in matching the right spots of

Cruise's face. Although the performance by both methods is very similar, the advanced matching methods seems to be more accurate as the simple matching methods finds a match on Cruise's right cheek which actually is not a match. Whereas the advanced matching method finds 9 matches of which all of them are correct, the simple matching method finds 10, out of which one match is faulty.



Figure 2: Matching results with a picture of Tom Cruise. Left: Simple matching method. Right: Advanced matching method.

### (c) Evaluate Rotation Invariance

To evaluate the capability of the matching methods to still find matches when images are rotated or tilted, I put a book on a table and took pictures from different angles. In order to compare the results on different angles, I used only the advanced RANSAC matching method. I was surprised to find that the RANSAC algorithm was able to find many matches for all angles and tilts. Even when whole images are rotated 90 degrees, it still finds many matches. An overview of found RANSAC matches for different angles is presented in figure 3. Although the salient point detector is able to find sufficient matches for all combinations of angles, the amount of matches found seems to decrease as the angle between the two pictures increases. This can be said since there are over 450 matches found between the front view and the 10 ° in-plane image, whereas the amount of found matches with a difference of 60 ° as shown in figure 3c is reduced to 280.

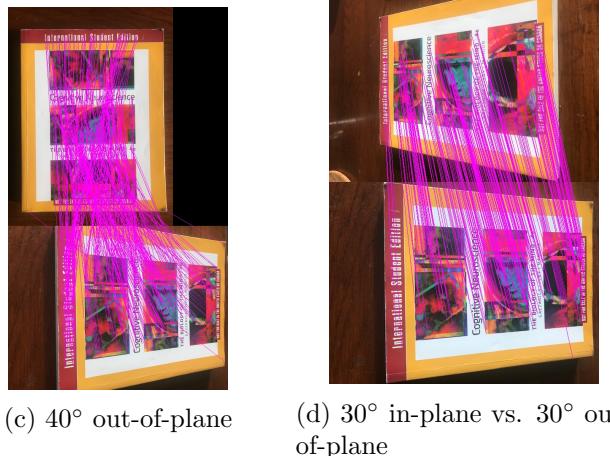
### (d) Which images are more similar?

In order to compare similarity of images, I slightly adjusted the main function of the getsift program, so that it requires three images as input, of which it calculates similarity between the first and the second image, as well as between the first and the third image. It is adjusted in such a way that it prints which of the images is more similar to the first image entered in the command line. The adjusted code with corresponding Make file is attached in the zip directory in which this journal is handed in. The program runs as presented in figure 4.



(a) 10° in-plane

(b) 40° in-plane



(c) 40° out-of-plane

(d) 30° in-plane vs. 30° out-of-plane

Figure 3: RANSAC matches for different angles

```
s1630784@huisu101:~/Documents/MIR02/get sift$ make
g++ -w alloc.cpp error.cpp kdtree.cpp sift.cpp xform.cpp cv.cpp test.cpp imgfeatures.cpp minpq.cpp smooth.cpp utils.
cpp -o similarimage -ljpeg -lm
s1630784@huisu101:~/Documents/MIR02/get sift$ ./similarimage front_view.jpg in_plane_10.jpg in_plane_40.jpg
SIFT Features Extraction: front_view.jpg
Numbers of Features from front_view.jpg: 967
SIFT Features Extraction: in_plane_10.jpg
Numbers of Features from in_plane_10.jpg: 836
SIFT Features Extraction: in_plane_40.jpg
Numbers of Features from in_plane_40.jpg: 844
coordinate and descriptor of front_view.jpg keypoints have been written in featfile1.txt
coordinate and descriptor of in_plane_10.jpg keypoints have been written in featfile2.txt
coordinate and descriptor of in_plane_40.jpg keypoints have been written in featfile3.txt
Found 502 total matches for in_plane_10.jpg
Found 462 total matches after RANSAC for in_plane_10.jpg
Found 264 total matches for in_plane_40.jpg
Found 280 total matches after RANSAC for in_plane_40.jpg
Based on the total amount of matches, in_plane_10.jpg is more similar to front_view.jpg than in_plane_40.jpg
s1630784@huisu101:~/Documents/MIR02/get sift$
```

Figure 4: Screenshot showing the execution and output of the adjusted getsift program.

These results as shown in figure 4 confirm the conclusion in task 1c, since the 10 ° in-plane picture is found to be more similar to the front view image of the book than the 40 ° in-plane picture.

## 2. Web Spider

The second task was to create a breadth-first web spider using the GetWebPage function and the GetLinksFromWebPage function as created in the second homework assignment. Breadth-first means that the webspider first crawls all weblinks on a superior webpage before crawling the weblinks on subpages.

In order to implement this, we were provided with pseudo-code as given in the lecture slides. The main goal was to crawl through webpages by adding found URL's into the back of a linked list in C. By using a linked list, it was possible to add the new URL's into the end of the list one by one, while deleting the first url from the list after parsing the URL's from it. These URL's were parsed from the html pages using the Haut html parser, which we were provided.

In addition to implementing this linked list we were supposed to implement some kind of tree structure in which the webpages that were already downloaded could be stored, so that they would not have to be downloaded again. Instead of using a regular B-tree, I flattened the binary tree in such a way that it could be handled as a linked list, just like the list in which new links are stored.

I succeeded in implementing a breadth-first web spider that crawls through webpages one by one. When starting at <https://www.universiteitleiden.nl/> it runs nicely until it reaches a webpage that can not be downloaded, which gives a segmentation fault. I expect that this is still due to a minor memory problem in the GetWebPage function as made in Homework assignment 2. So even though the web spider runs fine for as far I could judge, unfortunately it does not provide the expected results for 300 iterations. Figure 5 shows how the webspider is called, and what it does in the terminal. The output when provided with the URL <https://www.universiteitleiden.nl/> is included in figure 6, which shows a segmentation fault when trying to reach the site <https://www.universiteitleiden.nl/dossiers/444>. I could not yet find the exact reason why it gives the segmentation fault.

## 3. Google-lite: Inverted Index Search

Since I have been spending a huge amount of time and effort in building the web spider I did not yet have time to work on the third assignment, unfortunately. I will be catching up on this up as soon as possible.

(a) **Web Indices**

(b) **Webquery**

(c) **Image Search**

```
s1630784@huisuil03:~/Documents/MIR02/webspider$ ./webspider https://www.universiteitleiden.nl/
Downloading webpage:
https://www.universiteitleiden.nl/
HTML file downloaded successfully

Found the following weblinks: Adding them at the back of the queue...
https://www.universiteitleiden.nl/
https://www.student.universiteitleiden.nl/
https://www.medewerkers.universiteitleiden.nl/
https://www.organisatiegids.universiteitleiden.nl/
https://www.bibliotheek.universiteitleiden.nl/
https://www.universiteitleiden.nl/
https://www.universiteitleiden.nl/zoeken
https://www.universiteitleiden.nl/en
https://www.universiteitleiden.nl/
https://www.universiteitleiden.nl/onderzoek
https://www.universiteitleiden.nl/onderwijs
https://www.universiteitleiden.nl/wetenschappers
https://www.universiteitleiden.nl/over-ons
https://www.universiteitleiden.nl/over-ons/bestuur/faculteiten
https://www.universiteitleiden.nl/den-haag
http://www.bibliotheek.universiteitleiden.nl/
https://www.universiteitleiden.nl/dossiers/444
https://www.universiteitleiden.nl/nieuws/2019/03/lindsey-burggraaff-en-emma-koemans-winnen-voorrondे- famelab
https://www.universiteitleiden.nl/onderwijs/bachelors
https://www.universiteitleiden.nl/onderwijs/masters
https://www.universiteitleiden.nl/onderzoek/promoveren
https://www.universiteitleiden.nl/onderwijs/overig-onderwijs/onderwijs-voor-professionals
http://summerschool.universiteitleiden.nl/?utm_source=universiteitleiden&utm_medium=referral&utm_campaign=homepageheadernl
https://www.universiteitleiden.nl/agenda/extra/2019/03/proefstuderen-28-t-m-30-maart-2019
https://www.universiteitleiden.nl/agenda/extra/2019/03/masterdag
https://www.universiteitleiden.nl/nieuws/2019/03/leiden-krijgt-enorme-showcase-als-european-city-of-science
https://www.universiteitleiden.nl/nieuws/2019/03/neolithisch-huis-gaat-op-in-vlammen
https://www.universiteitleiden.nl/nieuws/2019/03/verslag-vrouwendag
https://www.universiteitleiden.nl/dossiers/actueel
https://www.universiteitleiden.nl/agenda/2019/03/this-weeks-discoveries---12-march-2019
https://www.universiteitleiden.nl/agenda/2019/03/mensenhandel-in-nederland
https://www.universiteitleiden.nl/agenda/2019/03/betabanenmarkt
https://www.universiteitleiden.nl/agenda
https://www.universiteitleiden.nl/onderwijs/bachelors
https://www.universiteitleiden.nl/onderwijs/masters
https://www.universiteitleiden.nl/onderzoek/promoveren
https://www.universiteitleiden.nl/onderwijs/overig-onderwijs/onderwijs-voor-professionals
http://summerschool.universiteitleiden.nl
https://www.universiteitleiden.nl/onderwijs/overig-onderwijs
https://www.universiteitleiden.nl/onderwijs/bachelors/voorlichtingsactiviteiten/open-dagen
https://www.universiteitleiden.nl/onderwijs/masters/voorlichtingsactiviteiten/masterdagen
https://www.universiteitleiden.nl/en/archaeology
https://www.universiteitleiden.nl/geesteswetenschappen
https://www.universiteitleiden.nl/geneeskunde-lumc
https://www.universiteitleiden.nl/governance-and-global-affairs
https://www.universiteitleiden.nl/rechtsgeleerdheid
```

Figure 5: Terminal output when running the web spider

```

Downloading webpage:
https://www.bibliotheek.universiteitleiden.nl/
HTML file downloaded successfully

Found the following weblinks: Adding them at the back of the queue...
https://www.universiteitleiden.nl/
https://www.student.universiteitleiden.nl/
https://www.medewerkers.universiteitleiden.nl/
https://www.organisatiegids.universiteitleiden.nl/
https://www.bibliotheek.universiteitleiden.nl/
https://www.bibliotheek.universiteitleiden.nl/
https://www.bibliotheek.universiteitleiden.nl/zoeken/onderwerp
https://www.bibliotheek.universiteitleiden.nl/zoeken/onderwerp
https://www.bibliotheek.universiteitleiden.nl/zoeken/onderwerp
https://www.bibliotheek.universiteitleiden.nl/zoeken/onderwerp
https://www.bibliotheek.universiteitleiden.nl/zoeken/onderwerp
https://www.bibliotheek.universiteitleiden.nl/zoeken/onderwerp
https://www.bibliotheek.universiteitleiden.nl/zoeken/onderwerp
https://www.library.universiteitleiden.nl/
https://www.bibliotheek.universiteitleiden.nl/
https://www.bibliotheek.universiteitleiden.nl/bezoek-en-gebruik
https://www.bibliotheek.universiteitleiden.nl/zeksystemen
https://www.bibliotheek.universiteitleiden.nl/training
https://www.bibliotheek.universiteitleiden.nl/onderzoek-en-publiceren
https://www.bibliotheek.universiteitleiden.nl/bijzondere-collecties
https://www.bibliotheek.universiteitleiden.nl/over-ons
https://www.bibliotheek.universiteitleiden.nl/steun-ons
https://catalogue.leidenuniv.nl/pds?func=load-login&calling_system=primo&institute=ubl&eng&url=https://catalogue.leidenuniv.nl:443/primo_library/libweb/id%3dubl_v1%26sortby%3drank%26lang%3den_us%26fromLogin%3dtrue%26from-new-ui%3d1%26authenticationprofile%3d31ukb_leu_primo_direct_pds
https://www.bibliotheek.universiteitleiden.nl/tel:071 527 2814
https://api.whatsapp.com/send?phone=31623108801
https://www.bibliotheek.universiteitleiden.nl/over-ons/contact/stel-een-vraag
https://www.bibliotheek.universiteitleiden.nl/nieuws/2019/03/rectoren-delen-eigen-werk-open-acces
https://www.bibliotheek.universiteitleiden.nl/nieuws/2019/02/grote-uitbreiding-collectie-eigenlijke-marokkaanse-literatuur
https://www.bibliotheek.universiteitleiden.nl/nieuws/2019/02/alumnus-schenkt-collectie-wenckebach-boekbanden
https://www.bibliotheek.universiteitleiden.nl/2019/02/docenten-maken-succesvol-gebruik-van-course-reserves-voor-collegeliteratuur-studenten
https://www.bibliotheek.universiteitleiden.nl/nieuws?category=library
https://www.bibliotheek.universiteitleiden.nl/mededelingen/2019/03/emerging-trends
https://www.bibliotheek.universiteitleiden.nl/mededelingen/2019/03/proefabonnement-op-askzad
https://www.bibliotheek.universiteitleiden.nl/mededelingen/2019/03/toegangsrestricties-universiteitsbibliotheek-9-maart-tot-en-met-5-april
https://www.bibliotheek.universiteitleiden.nl/mededelingen/2019/01/kijk-jezelf-vierkante-ogen
https://www.bibliotheek.universiteitleiden.nl/mededelingen?category=library
https://www.bibliotheek.universiteitleiden.nl/onderzoek-en-publiceren/centre-for-digital-scholarship
https://www.bibliotheek.universiteitleiden.nl/onderzoek-en-publiceren/open-access
https://www.bibliotheek.universiteitleiden.nl/nieuws/2019/01/tentoonstelling-minervas-metamorfosen-over-444-jaar-universiteit-leiden
https://studyspots.universiteitleiden.nl/
https://www.bibliotheek.universiteitleiden.nl/agenda/2019/01/minervas-metamorfosen-over-444-jaar-universiteit-leiden
https://www.bibliotheek.universiteitleiden.nl/agenda/2019/03/mensenhandel-in-nederland
https://www.bibliotheek.universiteitleiden.nl/agenda/2019/03/pelgrim-leven-en-reizen-van-christiaan-snouck-hurgronje
https://www.bibliotheek.universiteitleiden.nl/agenda/2019/03/universality-without-uniformity-early-childhood-parenting-across-the-globe
https://www.bibliotheek.universiteitleiden.nl/agenda
https://www.bibliotheek.universiteitleiden.nl/over-ons/locaties
https://www.bibliotheek.universiteitleiden.nl/over-ons/locaties/openingstijden
https://www.bibliotheek.universiteitleiden.nl/bezoek-en-gebruik/studeren-in-de-bibliotheek/vrije-computers
https://catalogue.leidenuniv.nl/primo_library/libweb/action/Login.do?loginfn=signin&id=ubl_v1&targeturl=http://catalogue.leidenuniv.nl/primo_library/li
https://www.bibliotheek.universiteitleiden.nl/bezoek-en-gebruik/thuistoegang
https://www.bibliotheek.universiteitleiden.nl/bezoek-en-gebruik/praktische-informatie-bijzondere-collecties
https://catalogue.leidenuniv.nl
https://digitalcollections.universiteitleiden.nl
https://databases.library.leiden.edu/?bibid=998024428790302711&redirect=true
https://databases.library.leiden.edu/?bibid=998009056650302711&redirect=true
https://scholar.google.com
https://www.bibliotheek.universiteitleiden.nl/over-ons/contact/stel-een-vraag
https://www.jotform3.leidenuniv.nl/form/73305826959
https://www.facebook.com/bleiden
https://twitter.com/bleiden
https://www.youtube.com/bleiden
https://www.organisatiegids.universiteitleiden.nl/reglementen/algemeen/universitaire-website-disclaimer
Could not download webpage
Segmentation fault (core dumped)

```

Figure 6: Webspider output results