A PIPELINE FOR AUTOMATIC PHENOTYPING AND MORPHOMETRIC ANALYSES OF PLANTS

PROJECT PROPOSAL

Mirna Baak

Bio-informatics at University of Applied Science Leiden

Student number: S1057641

Project duration: 02/09/2013 – 27/06/2014

Submission date: 4 -10 -2013

EC: 59

Supervisors of Naturalis Biodiversity Center:

Dr. Barbara Gravendeel and Dr. Rutger Vos

Naturalis Biodiversity Center
Sylviusweg 72
2333 BE Leiden

Supervisor of University of Applied Science Leiden:

Jan Oliehoek

University of Applied Science Leiden, Bio-informatics
Zernikedreef 11
2333 CK Leiden

CONTENTS

Introduction 3
Goal and Research Question 7
Products 8
Planning 9
Risk analysis 10
Project boundaries 11
Supervision 11
References 12
Appendices 13

INTRODUCTION

The need for new pipelines for automatic phenotyping

A lot of companies in the horticultural business want to separate seeds of weeds from their mixture of commercially sold seeds, crops or ornamentals. When there are less weed seeds in the mixture less weeds will grow between crops of interest. The efficiency of the horticultural land shall be higher and this is needed because of the current rising need for more agricultural products (Godfray, *et al.,* 2010).

Phenotypic information will change over time by the development of an individual, population or species in response to changing environmental conditions. Associated with genomic information, phenotypic variations could be linked with genetics. To link the phenotypic with the genetic information there is a need for high-quality phenotypic information. (Cobb, *et al.*,2013).

When phenotypes can be linked to genotypes seeds with useful genotypes can be separated for more efficient breeding. Also collection sorting can be done automatically with a pipeline for automatic phenotyping. When phenotyping can be done automatically it will save time and it will be less labour intensive.

Currently available software

An example of currently available software is the package 'SmartGrain' which was built for identifying rice seeds from a picture and compare these with each other so that morphological differences can be observed. This software is only for 2D pictures and not for 3D models. (Takanari, *et al*., 2012).

A fully automated measurement system that monitors the development of each individual seedling by high resolution time-lapse photography is developed at Wageningen University Research centre, The Netherlands by Kokorian *et al.* (Kokorian *et al.*, 2010). This system follows *Arabidopsis* seedlings in time with fixed cameras above the plants in such a way that changes in morphological traits of plant leaves can be monitored during development. Because this system uses highly sensitive and expensive fixed cameras it cannot measure outside in the field. The software written can only be used for the specific phenotype setup described by Kokorian *et al.* and cannot easily be adapted for other experiments. (Kokorian *et al.*,2010).

SPICY is a third technique developed for automatic phenotyping of large pepper plants in the greenhouse" (van der Heijden *et al.,* 2012). Here the cameras are not fixed. The plants are not transported to the camera but the camera is transported to the plants. This approach was chosen because peppers are too tall to transport. Stereo pictures are taken to make a 3D picture and the number of fruits on the plant can be calculated using statistics. (van der Heijden *et al.,* 2012). In this research 3D laser scanning is not used and a full scan of only the fruits is not possible.

Specific case study

This study will focus on the genus *Nepenthes* that encompasses tropical pitcher plants. All species are either carnivorous (majority) or herbivorous (minority) and very good in trapping insects or plant remains with specially modified leaf tips. *Nepenthes* is found in Africa and Asia (Jebb and Cheek. 2001). *Nepenthes rafflesiana* produces two different pitchers. The lower pitchers, that face the tendril, usually rest on the ground. A mature plant produces also upper pitchers that face away from the tendrils (Bauer, et al., 2010). The lower pitchers primarily feed on ground dwelling insects that enter the pitcher by walking over the wide wings with projections which act like 'ladders'. The upper pitchers feed on aerial insects attracted by other lures such as fragrances.

Figure 1: Nepenthes rafflesiana, upper pitcher(A) lower pitcher(B) (van Wely, 2013)

A previous study on the pitchers of *Nepenthes* was carried out by MSc student Valeri van Wely in 2013. In this study the ontogeny of *Nepenthes rafflesiana* pitchers was investigated. Due to severe trimming just months before the study of Valeri van Wely began, the *Nepenthes* plants were not able to produce sufficient numbers of mature upper pitchers. He 3D scanned pitchers from different developmental stages using the NextEngine 3D scanner of Naturalis Biodiversity Center. The 3D scanned pitchers were mainly lower pitchers. For my own internship research upper pitchers in different developing stages have to be scanned with the NextEngine 3D scanner as well. In the study of Valeri the raw 3D data was auto-aligned, trimmed and refined. In my own study the new raw 3D data of the scanned pitchers will also be modified. After scanning and modifying the 3D data, landmarks were applied with Landmark 3.0.0.6 software and some statistical analysis were done including a Principle Component Analysis. All these steps will be passed through for the newly scanned pitchers in my own project. In figure 2 a flow chart of the different steps of the project is presented.

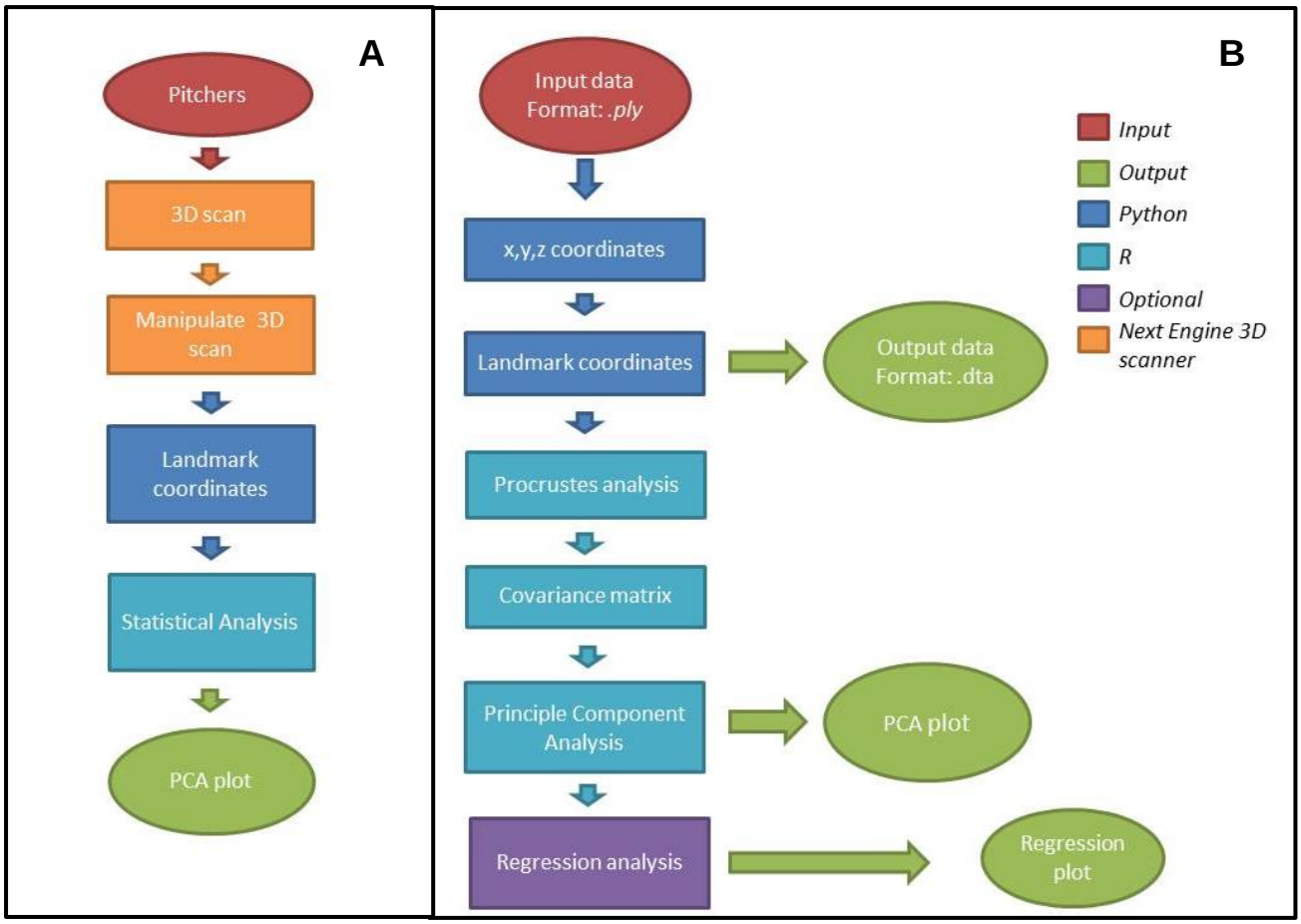

Figure 2: Flowchart of all steps to be carried out from scanning the pitchers up to producing a PCA plot (A) and analytical steps to be performed including the languages in which these will be written (B).

Once the pipeline has been built, its robustness will be tested by also analysing scans of two other species of *Nepenthes*. These species are *N. x hookeriana* and *N. ampullaria*. The latter species, *N. ampullaria*, is vegetarian and traps leaf litter with pitchers on the ground. The pitchers are clustered together above the soil surface (figure 3) and only lower pitchers are formed (Pavlovič, *et al.,* 2011).

Figure 3: N. ampullaria produces a cluster of pitchers on the ground for trapping leaf litter falling from the canopy above. (Pavlovič, et al., 2011)

N. x hookeriana is a natural hybrid between *N. rafflesiana* and *N. ampullaria*. The shapes of the *N. x hookeriana* pitchers, both the lower and upper ones, are a mixture of the shapes of the pitchers of *N. rafflesiana* and *N. ampullaria* (figure 4) (Yulita, *et al.* 2012). Hybrids between the carnivorous *N. rafflesiana* and vegetarian *N. ampullaria* are rare in the wild, possibly because the hybrid is neither good in digesting insects nor plant remains. *N. x hookeriana* has more morphometric similarities with *N. rafflesiana* so this species will be used as first test case for the robustness of the pipeline prior to the scanning and analysing of *N. ampullaria*.

Figure 4: Lower pitcher of Nepenthes x hookeriana (Yulita, et al. 2012).

GOAL AND RESEARCH QUESTION

The goal of this internship is to make a re-usable analytical pipeline for automatic phenotyping of plant organs. This pipeline will be written in R and python and should be capable of handling .ply data of 3D models and .dta data of landmarks to produce a principle component analysis plot.

The research questions are:

1. Are lower pitchers of *N. rafflesiana* primitive stadia of upper pitchers?

2. Are pitchers of *N. x hookeriana*, the hybrid between carnivorous *N. rafflesiana* and vegetarian *N. ampullaria*, maladaptive (i.e. neither good in digesting insects nor plants)?

If the first hypothesis is true, I expect the early developmental stages of both lower and upper pitchers of *N. rafflesiana* to overlap in the PCA plots. If this is not the case, this hypothesis should be rejected.

If the second hypothesis is true, I expect the pitchers of *N. x hookeriana* to fall in a separate, so-called maladaptive space in the PCA plots as compared with the pitchers of *N. rafflesiana* and *N. ampullaria.* If this is not the case, this hypothesis should be rejected.

Methods

Besides scanning separated pitchers, pitchers will be scanned during a restricted time period when still attached to the plant. The pitchers will be followed from the very beginning until it is clear that they are either lower or upper pitchers. This is needed because in the very beginning it is not easy to see yet whether they will develop into an upper or a lower pitcher. The 3D scanning protocol developed by Valeri will have to be adjusted for this purpose as I have to take the scanner to the Hortus botanicus and scan noninvasively (i.e. not 360 degrees).

A morphometric analysis of 3D data of the *Nepenthes* pitchers will be made and automatic phenotyping will be done. The x, y and z coordinates of the scanned pitchers will be converted to Principal Coordinates for statistical analysis.

The pipeline will accommodate all analytical steps previously carried out separately from the imaging onwards. As this will save a lot of time, so this pipeline will not only be useful for this particular case study. It should be possible to apply it to other projects as well focussing on seeds and fruits of crops and ornamentals. There is a huge demand for automatic phenotyping pipelines to link morphometric data to for instance genome data.

PRODUCTS

During this internship a number of products will be created. These will consist of presentations, a poster, reports and programming code.

Presentations
At the beginning of the project an introductory presentation will be done to present the goal of the project, the research question and methods.

During the internship another presentation will be given at the Character Evolution focus group of Naturalis Biodiversity Center and Kenniskringoverleg of the University of Applied Science Leiden. In these presentations the progress and preliminary results of the project will be discussed.

A final presentation about the entire research will be held at Naturalis Biodiversity Center for colleagues. This presentation has to be held before the graduation presentation.

The graduation presentation about the entire internship and the research is planned for week 40. This presentation will be held at the Naturalis Biodiversity Center in attendance of Barbara Gravendeel, Rutger Vos, Jan Oliehoek and a second examiner.

The presentation contains at least materials and methods, results, conclusion and discussion. The presentation will take about 20 to 30 minutes. After the presentation there will be questions about the presentation and the graduation report for 20 minutes. (Opleiding Bio-informatica, 2013)

Poster
A poster will be presented at the poster presentation of the University of Applied Science Leiden.

Report
The interim report has to be submitted in week 23 via Ephorus on ELO. It contains at least the following components:

- Introduction
- Material en Methods
- First Results
- A part of the Discussion for the graduation report.

A graduation report will be made of the results of the research. The report contains the following components:

- Abstract
- Introduction
- Material and Methods
- Results
- Conclusion and Discussion
- Reference
- Appendix

This report has to be submitted in week 37 via Ephorus on ELO and as four hardcopy reports to André Klein. (Opleiding Bio-informatica, 2013)

Programming code
The code written for the analytical pipeline shall be handed in on paper in annotated form. Other code that was written during the research will also be included at the report and digitally handed in on Github.

PLANNING

Task	**Internship week**
Writing project proposal	1-4
Literature study	1-2
Introduction to scanning	1-2
Introductory presentation	1-4
Scanning *Nepenthes pitchers*	2-25
First internship visit	2/3/4
Study statistical analyses	2-6
Morphometrics workshop	4 (25/9-27/9)
Return day 2 Mini training market	6
Analysis of 3D scans	2-25
Deadline Project proposal	8
Second internship visit	8/9/10
Study R ,Linux, git, Python	3-15
Programming statistical analyses	6-25
Programming analytical pipeline	6-34
Return day presentation production	11-12
Return day 3 Internship presentation	13
Interim Evaluation	16
Document analytical pipeline	25-35
Writing interim report	20-23
Deadline Interim Report	23
Third internship visit	23/24/25
Poster production	27-29
Return day 4 Poster presentation	30
Final report production	34-37
Presentation production	34-39
Deadline Final report	37
Presentation	36-40
Deadline Final presentation	40

RISK ANALYSIS

To minimize problems during the internship a risk analysis is made. A score is given for every risk. The score consist of the chance of the risk to occur, the impact and the discovery of the risk. The scores are multiplied and the different priorities of the risk can be observed. The higher the total score the higher the priority of the risk. The student has to alert her supervisors when risks with a high score develop during this internship.

Chance: 1(no chance or hardly) until 5 (in all probability)
Impact: 1 (no impact or hardly) until 5 (great impact)
Discover: 1 (easy to discover) until 5 (hardly to discover)

Risk	Chance	Impact	Discover	Total
3D scanner fail	1	4	1	4
No pitchers of *Nepenthes*	2	3	1	6
Supervisor absent for a while	1	3	2	6
Illness student	2	4	1	8
3D data lost	2	5	2	20
Software not finished at end of internship	2	5	2	20
Report not finished at end of internship	2	5	2	20
Wrong Interpretation of scans	2	5	3	30

Wrong interpretation of the 3D scans of the pitchers has a great impact. To minimize this risk, observations made have to be discussed with the supervisors. The supervisors can help to interpret these correctly.

When the software, which has to designed, is not finished at the end of the internship the internship cannot be finished correctly. To avoid this, a planning has to be made and followed. When the planning cannot be followed anymore the planning has to be reviewed and discussed with the supervisor.

For the report the same applies as for the software.

3D data lost is a risk that can be avoided by making backups of the files.

The student can be sometimes absent because of illness. When this will be for a long time, the supervisors have to be contacted on further action. Maybe the planning has to be reviewed to end the internship successfully.

The supervisor can be absent for a while. If this happens the student has to contact the supervisor to know who could be asked questions instead. For this internship there are two supervisors, so I expect no big problems.

Absence of pitchers of *Nepenthes* are also a risk. During a previous internship on *Nepenthes* the plants did not grow very well due to renovations of the greenhouse. These are over now, so we expect to collect more pitchers. Because there are already a couple of scanned pitchers available I can work with these in the meantime.

If the 3D scanner fails or is broken down there is a problem for scanning new pitchers. This risk is highly hypothetical as the scanner is brand new. If the 3D scanner breaks down nonetheless

and cannot be fixed a solution has to be devised with the supervisors. In Delft another machine is available and I could go there.

PROJECT BOUNDARIES

For this research there are a couple of boundaries. Genetic material of the pitchers will not be analysed. Only pitchers of *Nepenthes rafflesiana* , *Nepenthes x hookeriana* and *Nepenthes ampullaria* will be 3D scanned.

Only an analytical pipeline for *Nepenthes* pitchers will be developed. This pipeline has to be reusable for other research.

SUPERVISION

University of Applied Science Leiden:

Internship supervisor: Drs Jan Oliehoek

Jan Oliehoek is the internship supervisor of school. He supervised the student from University of Applied Science Leiden. He talks with the internship supervisor of Naturalis about the progress and the student during the internship. At the end of the internship he reviews the internship.

Naturalis Biodiversity Center:

Internship supervisor Dr. Barbara Gravendeel
Internship supervisor Dr. Rutger Vos

The supervisors of Naturalis will help the student during the internship and the student can ask questions related to the project.

REFERENCES

Bauer, U., Grafe, T. U. and Federle, W. 2010 Evidence for alternative trapping strategies in two forms of the pitcher plant, *Nepenthes rafflesiana. Journal of Experimental Botany,* 62: 3683-3692

Cobb, J.N., DeClerck, G., Greenberg, A., Clark, R. and McCouch, S. 2013. Review Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. *Springer Theoretic and Applied Genetics,* 126:867-887

Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M and Toulmin, C. 2010 Review Food Security: The Challenge of Feeding 9 Billion People. *Science*, 327:812-818

Heijden, G. van der, Song, Y., Horgan, G., Polder, G., Dieleman, A., Bink, M., Palloix, A., Eeuwijk, F. van and Glasbey, C. 2012. SPICY: towards automated phenotyping of large pepper plants in the greenhouse. *Functional Plant Biology,* 39: 870-877

Kokorian, J., Polder, G., Keurentjes, J.J.B., Vreugdenhil, D. and Olortegui Guzman, M. 2010. An ImageJ based measurement setup for automated phenotyping of plants. *ImageJ* 177-182

Opleiding Bio-informatica Hogeschool Leiden. 2013. Het afstudeerjaar 2013-2014 Opleiding Bio-informatica

Pavlovič, A., Slováková, L. and Šantrůček, J. 2011. Nutritional benefit from leaf litter utilization in the pitcher plant *Nepenthes ampullaria. Plant,Cel land Environment,* 34: 1865-1873

Takanari ,T., Taeko, S., Kiyosumi, H., Kaworu, E., and Masahiro, Y. 2012. SmartGrain High-Throughput Phenotyping Software for Measuring Seed Shape through Image Analysis. *Plant Physiology*, 160:1871-1880

Wely, V. van. 2013. Pitcher ontogeny of *Nepenthes* (Biology MSc thesis) *Naturalis Biodiversity Center*, 1-30

Yulita, K.S. and Mansur, M. 2012. The Occurrence of Hybrid in Nepenthes hookeriana Lindl. from Central Kalimantan can be Detected by RAPD and ISSR Markers. *Hayati Journal of Bioscience*, 19: 18-24

version 18-06-2013 by Valeri van Wely and 18-09-2013 by Mirna Baak

Protocol for the NextEngine 3D scanner HD

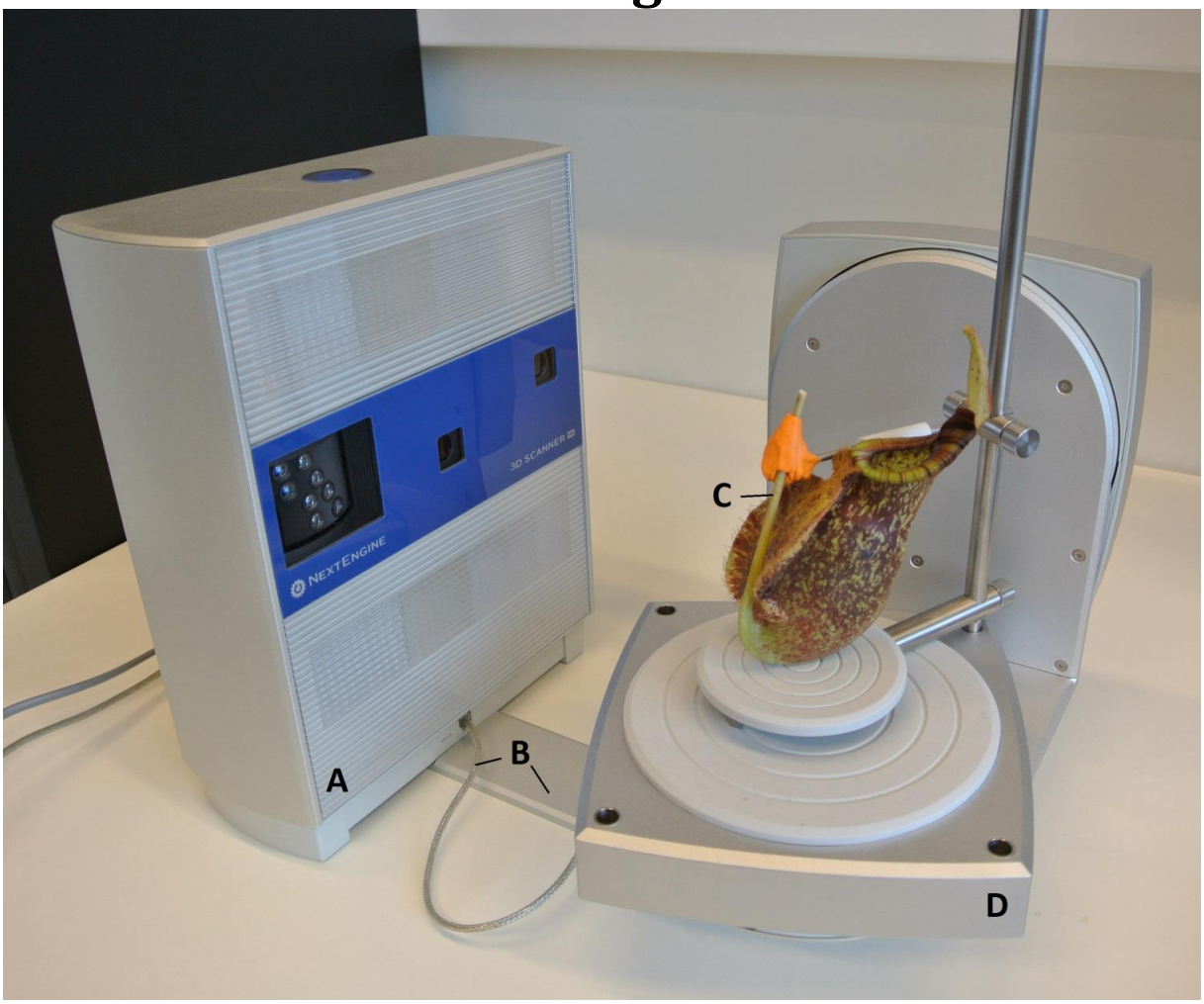

Figure 1: Assembled NextEngine 3D scanner HD. The 3D scanner (A) contains the lasers, lenses and camera's required for the 3D imaging. The object of interest (C) is mounted onto the tiltable MultiDrive (D), which is attached to 3D scanner with a cable and a metal connection (B). The object shown is around the maximum size for use with the MultiDrive. version 18-06-2013 by Valeri van Wely

Setting-up the 3D scanner

1) Assemble the 3D scanner to the 'MultiDrive' (= tilting table), see figure 1.

2) Start PC, connect USB cable and electrical plug.

3) Launch 'ScanStudio' software

4) Calibrate MultiDrive when reattaching: '*Align'* -> '*Calibrate MultiDrive'*

5) Press start (green arrow).

4) Attach specimen with the orange gum, a sticker or something else (make sure your object does not wobble when rotating!). Make sure the object is in the middle and will stay completely within the camera's view while rotating.

6) Select folder where you want to save your scans (and auto-backups): '*Edit*'-> '*Preferences'* (Do not use a Network folder, since this will drastically slow down the saving steps).

7) Select settings (see figure 2):

- **Positioning**: 360 for a complete 3D scan (Brackets=3 scans and Single scan=1 scan)
- **Divisions**: 16 divisions is maximal and most accurate. (Less decreases scanning time and quality).
- **Tilt**: Can tilt the table prior to 3D scanning.
- **Points/Inch**: 40k (highest resolution).
- **Target**: Neutral (unless either completely white or pitch-black)

7) (Optional) Because scanning 360 degrees around the specimen does not always scan the extreme lower or top part of the object, you might consider adding an extra '**Scan family**' which scans the object from another angle by tilting the MultiDrive. To do so, click the dot at tab 'B' and select slightly different settings for the second scan: 4 divisions from tilted angle is often enough (Single scan also possible), select tilt value that shows the otherwise unscanned part of object. After scanning, the ScanStudio software can automatically recognize and align the scans from the selected families, as long as the MultiDrive is used. An extra scanfamily can also be run after the first one is finished. The time at the bottom of the screen indicates the total time of all activated scan families together quite well.

7) **Noninvasive scanning (not 360 degrees)**
For noninvasive scanning use the option single or bracket scan at the scan menu. Please note: automatically align has to be set off, otherwise the turning table will rotate 360 degrees.
'Scan' > 'Settings' > Disable Scan- time AutoAlignment

- **Bracket scan:** The number of division is correlated to the amount of rotation degrees. So if division is set to 4, the object will rotate 90 degrees clockwise and counterclockwise from the center scan. And if the division is set to 16, (360/16) 22.5 degrees. The horizontal angle can be defined with the option 'tilt'
- **Single scan:** One scan will be performed. With the option start, the position can be set.

For scanning the entire object, multiple scans have to be performed. Because autoalignment is set off after scanning, an alignment have to be done by hand.

8) Press scan and wait until finished (don't move the scanner or the table while scanning). Lights can stay on during scanning, but strong sunlight opposing the camera's view should be avoided (this can be simply be done by turning the 3D scanner prior to scanning).

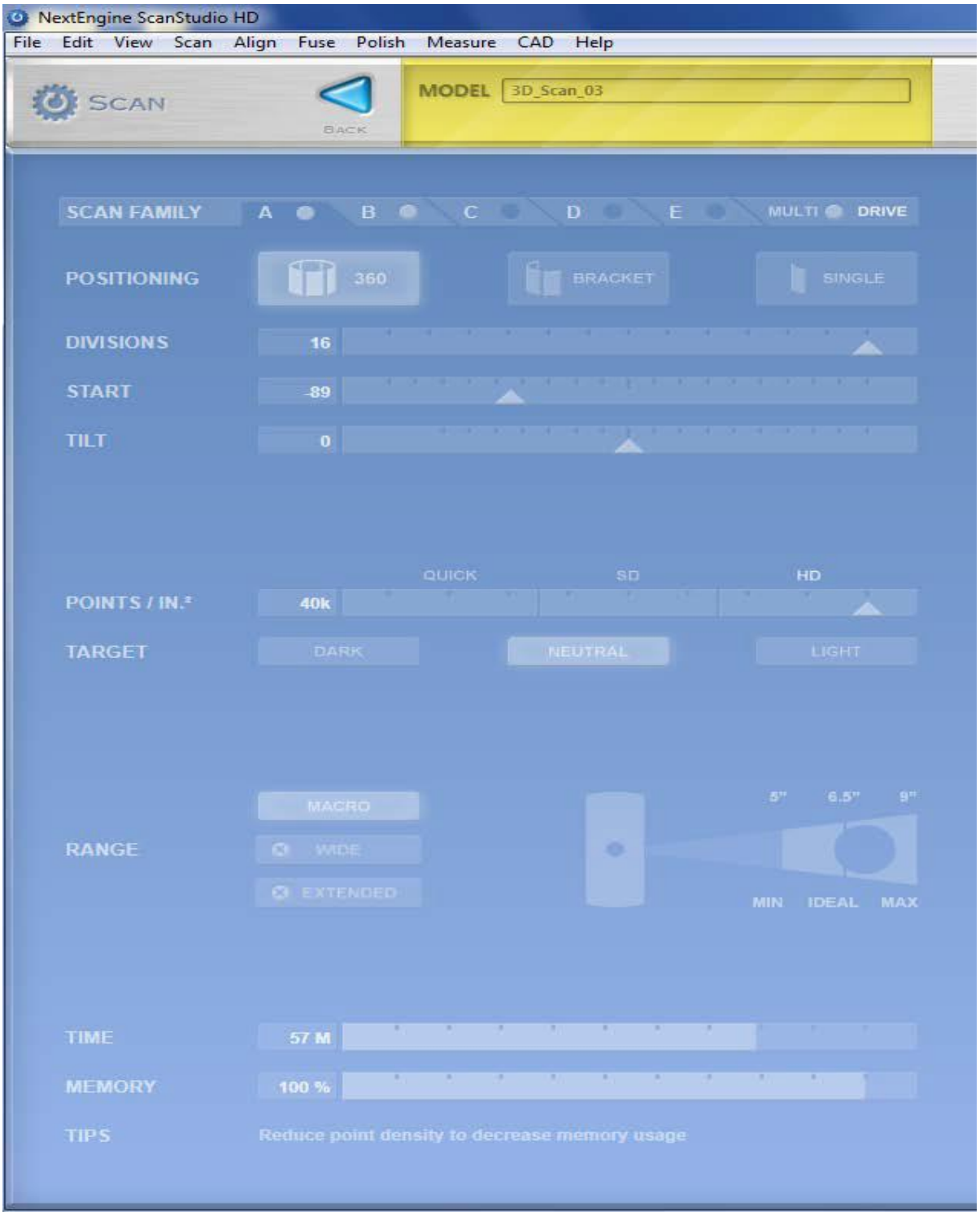

Figure 2: Possible settings for a 3D scan *2 scan families are activated in this example. The one shown here is family A, with the following settings: 16 divisions, no tilt, highest resolution.* version 18-06- 2013 by Valeri van Wely

Processing data after 3D scanning

A quick automated alignment will be performed automatically after scanning when using the MultiDrive and invasive scanning.

9) **Align for non-invasive scanning**
Click the Align button at the top of the screen. The screen will split in two separate screens. Drag the three colour points to different points on the model at the left. Place the three colour points at exactly the same location on the right model. More points can be chosen if needed. When done press 'Align' to perform the alignment. Repeat this step until all scans have been merged together.

10) **Trimming:** Press "Trim" and select area's you want to remove (such as the specimen holder or the rotation table). Wrongly selected areas can be 'unselected' again by using the '**-**' button. Press 'Trim' to remove selected areas. When ready, press 'Back'.

11) **Refining**: Press 'Align'->'Refine'. This often increases the quality of the alignment compared to the initial (quick) alignment.

12) Save model in the format you need. (.stl is a common format and .ply is commonly used in the landmark analysis software. A .scn file can be opened in ScanStudio again).

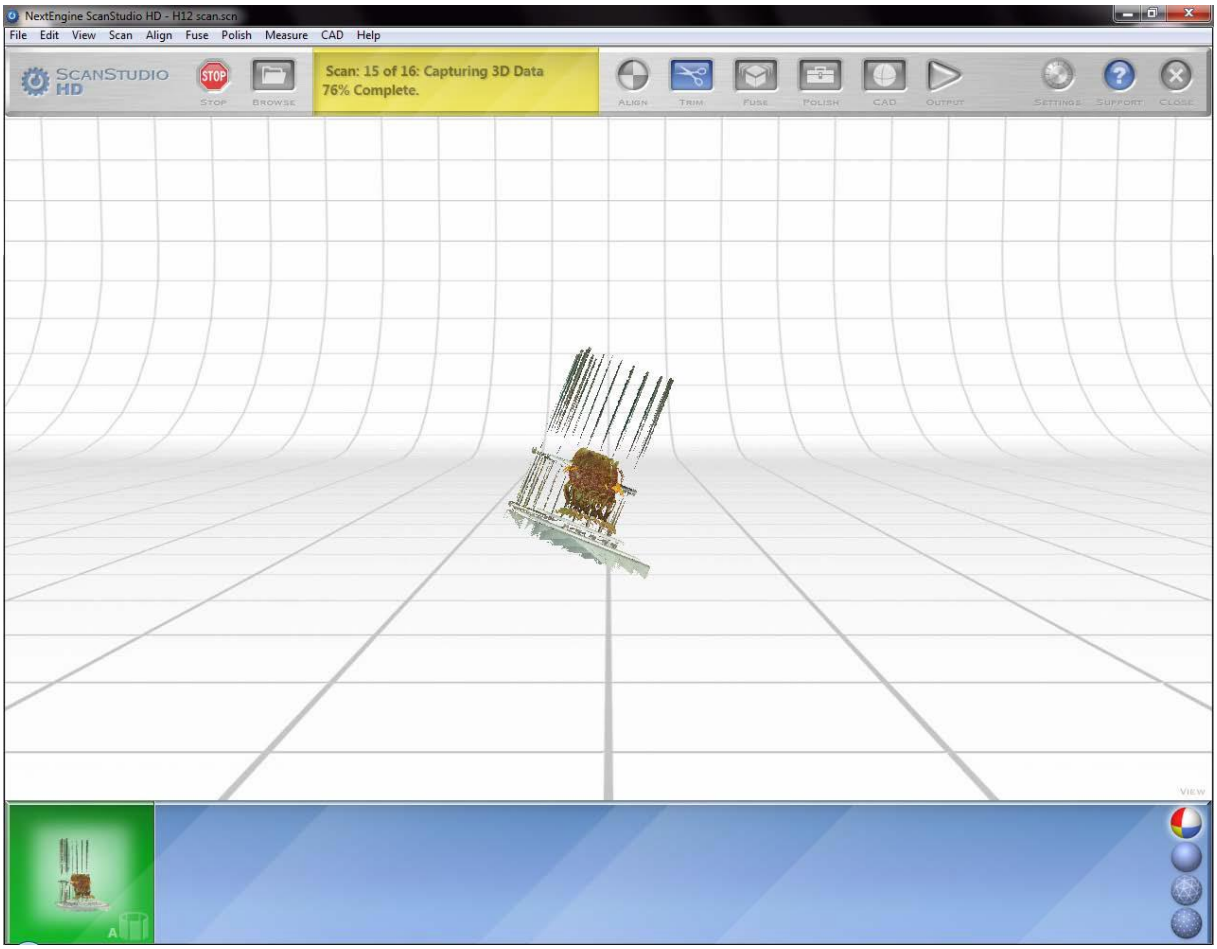

Figure 3: Screen that appears during and after scanning. *Several options are shown on top (trimming, aligning (manually) and saving ('output')), as well as a poorly aligned imaged in the middle. When scanning is complete, an automated alignment will be performed* version 18-06-2013 by Valeri van Wely

Additional Information:

• Please note: The 3D scanner is connected to the computer via a USB. Removing other USB devices (e.g. a USB stick) somehow leads to an error in the ScanStudio software, after which you have to do your scan again. In short: Do not remove USB devices while scanning!

• The MultiDrive can be attached to the 3D scanner with a screw at the bottom of the metal connection (B in figure 1), this provides extra stability for the scanner. However, if the scanner needs the be broken down after you finished scanning, it might be better to leave the screw, since it is a bit tricky to attach the screw while holding the scanner and the MultiDrive.

• The MultiDrive is always at a fixed distance from the 3D scanner and this greatly eases the automated alignment of the 3D scans by the ScanStudio software. However, because of this fixed distance, you can only use the MultiDrive in 'Macromode'. For bigger scans that don't fit in macromode (> +/-150 cm in length) you can use the other platform (the 'AutoDrive', which can't tilt), though aligning is less easy when using this platform. Aligning with the AutoDrive can be done manually in the ScanStudio software but can take you about an extra hour. So use the MultiDrive when possible for the quickest and the best alignment results.