EXTRACTOR

Read our article:

Dupej, J., Lacoste Jeanson, A., Pelikán, J., & Brůžek, J. (2017). Semiautomatic extraction of cortical thickness and diaphyseal curvature from CT scans. *American Journal of Physical Anthropology*, 164(4), 868–876. http://doi.org/10.1002/ajpa.23315

Data files formatting

The Extractor automatically extracts diaphyseal data from CT scans saved under RAW format. On top of the RAW files, it needs a single CSV datafile containing specification per specimen (one raw per specimen).

Prepare a CSV spreadsheet with the following information put column by column per specimen:

A) name of the RAW file with its extension (e.g. "indiv01.raw");
B) slice width (in pixels);
C) slice height (in pixels);
D) number of slices;
E) voxel format is either "u16", "i16", "i16b", "f32", "f32b" ("u" is unsigned integer; "i" is signed integer; "f" is float; "b" is big endian–if not present, little endian is assumed);
F) voxel x dimensions (in mm);
G) voxel y dimension (in mm);
H) voxel z dimension (in mm);
I) this value is the Houndsfield unit offset which is used to convert gray levels and is added to each voxel upon load–typically "-1000";
J) proximal landmark x (as described in Dupej et al., *American Journal of Physical Anthropology*, 2017);
K) proximal landmark y;
L) proximal landmark z;
M) distal landmark x;
N) distal landmark y;
O) distal landmark z;
P) orientation landmark x;
Q) orientation landmark y;
R) orientation landmark z;
S) future file name of the cortical thickness measures–**it should be set as "*_thickness.csv" ("*" being the name of the specimen)**;
T) future file name for the sampled axis–**it should be set as "*_axis.csv" ("*" being the name of the specimen)**;
U) future file name for the areas measured on each cross-section (CA)–**it should be set as "*_CA.csv" ("*" being the name of the specimen)**.

A model is provided at https://lite.framacalc.org/extractor_model.

The CSV file has to be upload under the button *Load batch*.

Software settings

The Extractor is an EXE file. The button "Load batch" should be used to load the CSV file created under the above-described conditions. Before running the extraction (button "Run"), a few settings can be adjusted by the user in the right panel. One the following parameters are set, the user just has to click on the *Engage!* button and wait for the extraction to be automatically proceeded.

AxisDetection

This setting specifies the method of identifying the medial axis.

Landmarks

Given the landmarks p_0, p_1, the medial axis is specified as the straight line between them.

$$\mathbf{a}_i = \mathbf{p}_0 + (\mathbf{p}_1 - \mathbf{p}_0)\left(r_0 + \frac{(i-1)(r_1 - r_0)}{n-1}\right); i = 1, ..n$$

CrossSectionCentroids

This setting forces the initial detection of medial axis as in the equation above. However, these axis points are then refined as follows. A plane perpendicular to the landmark- defined medial axis is constructed in each a_i from the previous step. Rough segmentation of cortical area is performed by thresholding against h_{prior}, in each reslice. Continuous elements of the segmented cortical area are determined and all but the largest one (by pixel count) are discarded. The centroid of the remaining cortical area in the reslice is calculated and stored as a_i.

CrossSectionLinear

Cross-section centroids are identified as with the setting *CrossSectionCentroids*. To these centroids, a straight line $A : c + td, t \in R$ is fitted with ordinary least squares. Points $c_0, c_1 \in A$ are found, such that the slice numbers of these match the slice numbers corresponding to the start and end of medial axis ROI. The medial axis $a_i, i = 1, ..n$ is then constructed by equidistantly placing points on the straight line between c_0 and c_1.

CrossSectionGaussian

This is the default setting. A preliminary axis $\tilde{a}_i$ is constructed the same as in *CrossSectionCentroids*. The final medial axis is established by convolving with a Gaussian kernel $a = \tilde{a} * k$, where

$$k(x) = exp\left(-\frac{x^2}{2\sigma_a^2}\right)$$

This setting is recommended for most applications. The medial axis is represented as a general polyline; however, the high-frequency information that can be caused by noise is suppressed. The width of the kernel σa, is configured in another field.

AxisGaussianBandwidth

Symbol σ_a
Valid range $(0,\infty)$
Default 3

This parameter is applicable only when axis detection is set to CrossSectionGaussian; with other settings, it is ignored. It controls the bandwidth of the Gaussian kernel that is used to smoothen the general medial axis. This value is not given in any length units, but in indices of the medial axis samples. Larger values will result in a smoother axis, very large values will restrict the responsiveness of the medial axis shape to the shape of the bone. Low values will cause the medial axis to better follow the shape of the bone; however very low values may not suppress the effect of noise in the data.

BoneROIStart

Symbol r_0
Valid range $(0,1)$
Default 0.2

This value specifies the location along the length of the bone, where reslicing starts.

BoneROIEnd

Symbol r1
Valid range $(0,1)$
Default 0.8

This value specifies the location along the length of the bone, where reslicing ends.

DebugAll

Specifies, if the result of each extraction is shown for diagnostic purposes. Note that is this is set to True, a debugging window will be shown after each extraction and the program proceeds with the next file only after the user closes the debug window. The result files are generated, regardless how this field is set.

MaxBoneRadius

Symbol k_{median}
Valid range $(60,100)$
Default 60

This value corresponds to the window (in pixels) around the bone that the extractor will take into account from the centroids–the default value is 60 and you can set it up to 100 in case bones are very robust. If too wide, the extractor will be less efficient and might carry some noise, if any.

MedianFilterRadius

This value configures the size of the median filter applied to the entire stack before extraction. Note that the median filtering is performed in 3D. Higher values will result in smother images, at considerable computational expense. If set to 0, no filtering is done.

NumRays

Symbol n_{ray}
Valid range ⟨1,10000⟩
Default 50

Sets the number of of rays the extractor will analyze from the centroid of the cross-section per cross-section. This is also equal to the number of columns of the cortical thickness matrix that results from the extraction.

NumSlices

Symbol n_{slice}
Valid range ⟨1,10000⟩
Default 100

Sets the number of cross-section reslices that will be produced in the extraction process within selected ROIs. This is also equal to the number of rows of the cortical thickness matrix that results from the extraction.

PriorThreshold

Symbol h_{prior}
Valid range ⟨−1000,3092⟩
Default 525

This is the prior threshold separating bone from soft tissues and is given in Hounsfield units. This value is used in rough segmentation of cortical area, and in some parts of the extraction, may be automatically refined.

RefineEndpoints

If this is set to *True*, endpoint refinement, as detailed in Dupej et al. (American Journal of Physical Anthropology, 2017) will be applied. It is recommended to enable this only when significant error in endpoint landmarks is expected. The default value is *False*.

ResultPath

Defines the name of the file where the results of the extraction will be stored.

UseHMH

If set to *True,* half-maximum-height adaptive thresholding (see Spoor et al., *American Journal of Physical Anthropology*, 1993) will be used. This will result in sub-pixel resolution in the detection of outer and inner cortical boundaries. If set to *False*, ordinary thresholding will be used. The default is *True* and it is recommended to keep this setting.

Extracted files

The semi-automatic extraction results in the creation of five files per specimen. The results are contained in the file for which the user has defined a name at the *ResultPath* setting in the right panel.

*_thickness.csv

Gives the thickness in mm of the entire diaphyseal portion selected: rows correspond to the horizontal plane of the diaphysis (i.e. to the cross-sections, corresponding to the function *NumSlices* and set by default *n*=100); columns correspond to the vertical plane (anterior, medial, lateral and posterior planes, corresponding to the function *NumRays* and set by default *n*=50).

*_axis.csv

Provides three dimensional measures of the medial axis of the bone at every cross-section of the considered potion of the diaphysis (corresponding to the function *NumSlices* and set by default *n*=100).

In order to get measures of the entire radius of the medial axis per individual, use R and follow the steps:

1. Put CSV files of the sampled axis in the directory,
2. Use the R script available at https://annuel.framapad.org/p/extractor_get_radius_medial_axis
3. If used
    - on a single file, type after the script: summary(fit.circle(filename="….csv"))
    - on multiple files type after the script: sapply(dir(pattern="*.csv"), function(t) fit.circle(filename=t)$r)
4. The results appear in R, by alphabetical order of the files.

*_CA.csv

First column provides medullary areas (inner surface) in mm^2 starting from the most proximal part of the diaphysis (the last row being the most distal part).

Second column gives total areas of the entire cross-sections (outer surface) starting from the most proximal part of the diaphysis (the last row being the most distal part).

*_axis.csv.inner.obj & _axis.csv.outer.obj

Inner file matches with the inner surface of the bone, i.e. the endosteal surface and the medullary cavity.

Outer file corresponds to the outer surface of the bone, namely the periosteal surface or external surface of the bone.