ARFS2Batch Help Document

Created 2016.08.22 by Alex Salmon
Last Updated 2016.08.22 by Alex Salmon
Current Version of ARFS algorithm: 1.1.1
Current Version of Batch Interface: 1.2.2
Current Contact: asalmon@mcw.edu, 1-414-955-2010

Section 1: Understanding ARFS

ARFS stands for Automated reference frame selection, and was developed to (1) save human researchers' time and energy, and (2) introduce objectivity in frame selection. ARFS can be broken down into two main modules: (1) distortion measurement, and (2) motion tracking.

Section 1.1: Distortion

Since we are using a scanning instrument on a moving object, it is guaranteed that there will be regions of the images that are over- or under-sampled by the system, i.e., some retinal structures are imaged more or less than they should be, or not at all. Fortunately, most retinas have limited variability in their organization, and major changes to their appearance can be detected with image processing techniques. ARFS takes advantage of the fact that the photoreceptors are usually circular and are tightly packed together (Fig. 1A). When one computes the discrete Fourier transform (DFT) of an image of photoreceptors, Yellott's ring, (Yellott, 1982) an annulus which indicates the modal spacing between photoreceptors, may be observed (Fig. 1B, C). Eye motion and subsequent oversampling of cones disrupts the circularity of Yellott's ring (Fig. 1D).
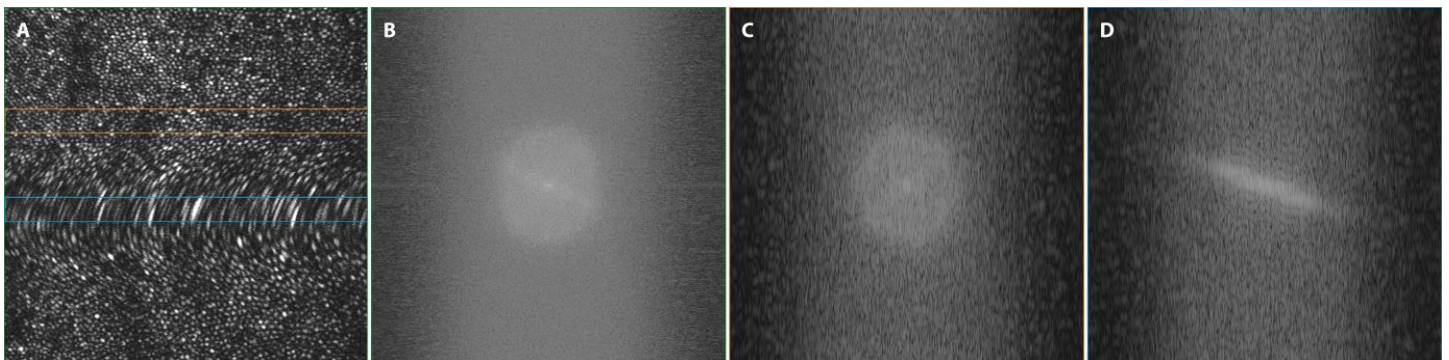

Fig. 1 Distortion detection. (A) A raw frame of an AOSLO image sequence with an obvious eye motion artifact. (B) Discrete Fourier transform (DFT) of A. (C) DFT of orange strip in A. (D) DFT of cyan strip in A. Note the disturbance to the circularity of Yellott's ring.

ARFS breaks each frame of an AOSLO video into strips and computes the DFT of each strip. The DFT is then masked to exclude anything outside the approximate area of Yellott's ring, and thresholded to include only significant contributors to the signal. The DFT is then binarized to include only the maximum value in each column and a line is fit using least-squares regression to the coordinates of the maxima. The R^2 value from a regression of an undistorted strip will be close to 0, whereas a distorted strip will result in a higher R^2 value. The R^2's from each strip are then summed to give a score for the whole frame. The frame with the lowest R^2 value at the end of processing is deemed the least-distorted frame.

Section 1.2: Motion Tracking

Anyone who has participated as a subject in AOSLO imaging can attest that perfectly fixating on a point of light isn't quite as easy as it sounds, and for subjects with pathological nystagmus, impossible. Eyes drift, correct, and repeat, involuntarily. When processing AOSLO images, it is important to build a mental model of this eye motion, so as to select a spatially representative frame. However, with our current methods, this is subjective, and at times physically taxing. For this reason, we developed ARFS to track eye motion between frames and look across the image sequence to group frames into spatial clusters. A normalized cross-correlation (NCC) matrix is calculated for a frame and the next frame in the sequence, and so on until the end is reached. The NCC is masked to exclude edge-artifacts. The height of the peak of the NCC matrix indicates how well two images overlap, and the position of the peak indicates how to position the two images such that they overlap the best (which is an indication of how far the eye moved between these two images). ARFS uses a custom peak-finding method which is best suited for NCC's of confocal images of cone photoreceptors. If a peak can be found, the relative positions of the two frames are recorded. If not, no new positions are recorded, and this point in the sequence is defined as the boundary between two groups of frames. After all individual frames are compared to their neighbors, whole groups of

frames may be moved if they are found to overlap. This is determined by calculating the NCC between the first frames in the two groups, the last frames in the two groups, as well as the first frame in one to the frame closest to the centroid of the other group. The rationale for doing these comparisons is based on the common event where a subject's eye drifts from the target, then quickly saccades back to the target. Frames for which a peak is never found are rejected from consideration as a reference frame. If a group cannot be registered to any other group, it is processed as a separate cluster with no known spatial relationship to the rest of the image sequence. Additionally, the distribution of NCC peak heights is examined, and frames with NCC peak heights less than the mean (μ) minus three standard deviations (σ) are rejected.

The frames are then sorted into spatial groups using the *k*-means++ (Arthur & Vassilvitskii, 2007) algorithm employed by Matlab. The number of clusters to start with is determined by the silhouette statistic, (Rousseeuw, 1987) which is a ratio of inter- to intra-cluster variance. This doesn't take into account the size of the frames, and may result in over-clustering, which would be deleterious as we would like to output the least-distorted frame from each distinct spatial cluster, and over-clustering may result in outputting a distorted image from a frame that is not significantly separated from a less distorted frame. To correct for this, frames may be absorbed if they would overlap with another cluster centroid by over 75%. This proceeds by decreasing cluster size. At this point, and after every subsequent frame rejection, it is assessed whether a cluster has more frames than a user-defined threshold for the minimum number of frames that may constitute a cluster. The thinking behind this threshold is to only process frames from clusters with enough frames to significantly increase the signal-to-noise ratio (SNR). Given an average quality dataset, if a cluster contains only four frames, for example, registering and averaging four frames is unlikely to increase the image quality. All of these frames would then be rejected. Additionally, distance to the cluster centroid is calculated for each frame, and outliers are rejected.

Section 1.3: Algorithm Process

1. Reject dark frames (mean pixel intensity $< μ - 3σ$)
2. Reject frames with significant distortions ($ΣR^2 > μ + 3σ$)
3. Conduct an additional strict rejection with both mean pixel intensity and distortion ($2σ$)
4. Get 1st pass of motion trace (compare individual frames)
5. Get 2nd pass of motion trace (compare groups of frames)
6. Reject frames that don't overlap (no peak found with this frame and any other group)
7. Reject frames with low NCC peaks (NCC $< μ - 3σ$)
8. Cluster frames
9. Reject frames too far from the cluster centroid (distance $> μ + 3σ$)
10. Reject small clusters
11. Strict NCC rejection (20th percentile and lower)
12. Output frames with lowest $ΣR^2$ in each cluster

Section 2: ARFS2Batch

ARFS2Batch is an environment in which to run ARFS on multiple videos. It was named as such because it acts as a bridge between ARFS and Create & Modify Batch, which creates dmb (DeMotion Batch) files used as inputs for strip-registration (Dubra & Harvey, 2010). Currently, it requires dmb files as inputs, and creates new dmb files with updated video names and reference frames.

Section 2.1: Pre-requisites

There are a few steps to complete before running ARFS2Batch. These include:
1. Ensure file names and folder structures conform to AOIP standard
2. Creating a desinusoid file
3. Creating secondary sequences
4. Establishing DeMotion sets

Further, ARFS2Batch only works on uncompressed monochrome avi's.

Section 2.1.1: File names and folder structure

ARFS2Batch attempts to do as much heavy lifting as possible, which requires a fair bit of standardization in folder structure and file naming in order to work well. We are adding fail safes as they present themselves, but best practice is to adhere to the following principles.

File names for videos and Savior data files should always be of the form: Ref_ID_λnm_OX_modality_vidnum, where Ref is a two-character token signifying the referring PI, ID is a unique number (leading 0's okay; no letters or special characters), λ is a number signifying the imaging wavelength used followed by nm, OX indicates the eye (either OS or OD), the modality may be 'confocal', 'split_det', 'avg', etc., and in future releases could be anything (as long as it's used consistently throughout

the dataset), but may currently crash if it encounters and modality I've never heard of. It is especially important that the video number is the last token before the file extension, and that the modality immediately precedes the video number. ARFS2Batch has a built in field-of-view (FOV) extractor which should make it easier for establishing DeMotion sets (below). For this component to work properly, it is best to have a root folder (name is unimportant) containing a Raw and Calibration folder. The Raw folder should contain all dmb's, videos, and Savior data files to be processed. The Calibration folder should contain the desinusoid file (see below), the original videos of the grids and their associated Savior data files.

File naming:

Example of acceptable filename: JC_10549_790nm_OD_confocal_0000.avi

Video number should be last element before file extension: JC_10549_790nm_OD_confocal_0000_anything.avi

Modality should always precede video number: JC_10549_790nm_confocal_OD_0000.avi

Folder structure:

All desinusoid files should be here, including the original grid .avi's and their associated .mat's.
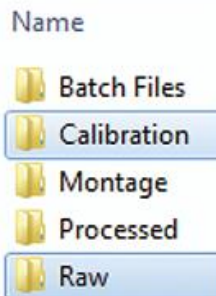
.dmb's, .mat's, and .avi's should all be here.

Name
- Batch Files
- Calibration
- Montage
- Processed
- Raw

Fig. 2 File names and folder structure. Best practices for automated dmb generation.

If the original grid videos are not available, ARFS2Batch may still infer the FOV from the name of the desinusoid file. For example, ARFS2Batch will scan through the parts of this file name:

    desinusoid_matrix_790nm_1p00_deg_118p1_lpmm_fringe_14p882_pix.mat

and determine whether only one token can be converted to a number by replacing a 'p' with a decimal (this would fail 1p00, 118p1, and 14p882 all work), whether 'deg' is part of a token that also contains a number (this would fail because 'deg' is all alone), or whether 'deg' is preceded by a token which can be converted to a number (this would succeed). If all else fails, the user will be asked what FOV belongs to what desinusoid file, which should happen very soon after the initial steps of setting up ARFS2Batch so no one returns to their terminal in the morning to find out it couldn't finish because it couldn't determine the FOV.

Section 2.1.2: Desinusoiding

Before running ARFS2Batch it is necessary to correct for the sinusoidal distortion caused by scanning (desinusoiding). This is done by imaging grids with known spacing and resampling the videos. In our lab, the script for estimating and correcting sinusoidal distortion can be found here (run roc_create_desinusoid_matrix_3p3.m).

Section 2.1.3: Secondary Sequences

If you would like to include secondary sequences (e.g., split and avg), you must generate them prior to running ARFS2Batch. In our lab, the script to generate these files can be found here (run Create_split_and_or_avg_movies_20160212_AES.m).

Section 2.1.4: DeMotion Sets

It is then necessary to formulate a processing plan. For certain regions of retina, one might use the same parameters for strip-registration, and only change the videos to process and the reference frames to use; this defines what I call a "DeMotion set" (DeMotion is the name of our strip-registration program). It is necessary to generate a template dmb for each DeMotion set. The minimum number of DeMotion sets is equal to the number of FOV's present in the dataset to be processed. For example, one of our standard protocols involves 1° and 1.5° FOV images at the fovea, as well as 2° FOV images between 1-10T and 1-10S. All the parameters are likely to be the same for the 1° FOV images, so 1 dmb needs to be made for this DeMotion set. The 1.5° FOV images constitute another DeMotion set since they require a different desinusoid file than the 1° FOV images, but all the strip-registration parameters could be maintained. Since cone diameter increases with eccentricity, the 2° FOV images may require different parameters after a certain eccentricity. The point at which one might want to increase the strip-size could be determined for the temporal and superior strip and divide them into DeMotion sets; let's say at 3T and 4S it becomes necessary to increase the strip-size, 1-2T and 1-3S are one DeMotion set, and 3T-10T

and 4S-10S are another. Thus, for this dataset we have 4 DeMotion sets, and have to generate 4 template dmbs with the correct desinusoid files, secondary sequence file names, and registration parameters (Fig. 2).
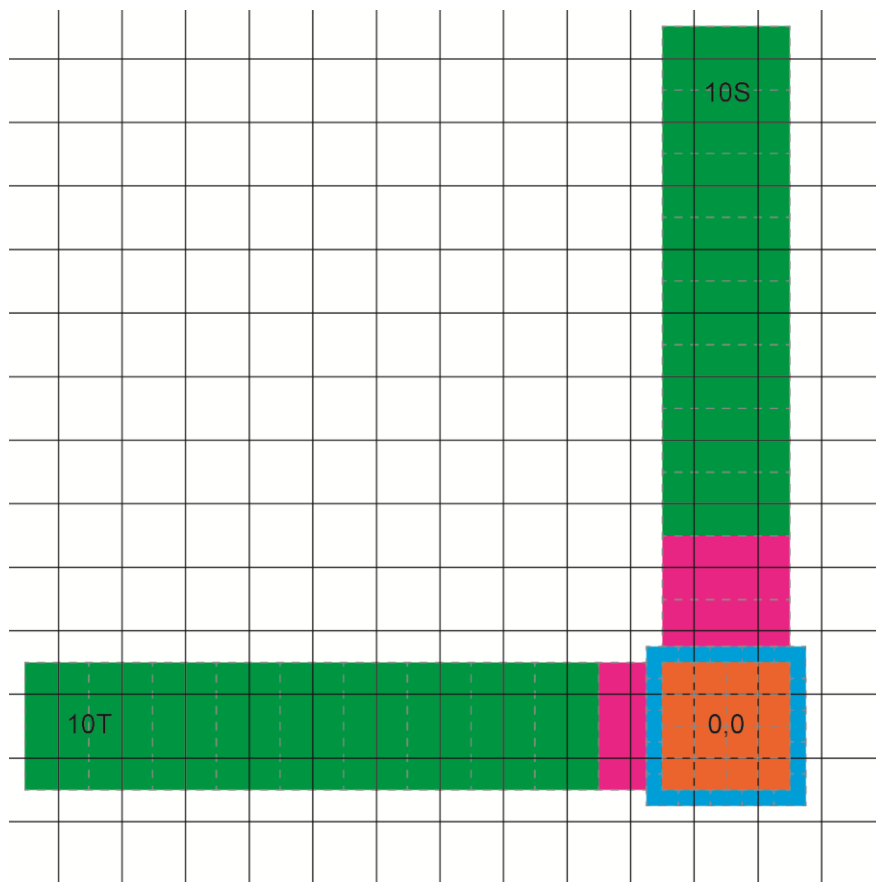

Fig. 3 Common protocol divided into DeMotion sets. 4 DeMotion sets are probably required for this protocol. 1° FOV images about the fovea (orange; 1), 1.5° FOV images about the fovea (cyan; 2), 2° FOV images at 1-2T and 1-3S (magenta; 3), and 3-10T and 4-10S (green; 4).

This process does require examining each video individually and estimating feature size and determining which modality would be the best primary sequence. If you have time (running over night, for example), you may want to double the number of template dmbs by choosing confocal as the primary for one group of dmbs and split or average as the primary for another. Keep in mind this means that all the calculations will be done on all the videos twice, but it may save time in the long run without having to switch around primaries and secondaries, especially in large datasets. You may find it useful to organize all videos with prominent blood vessels into a DeMotion set.

Section 2.2: Running ARFS2Batch

After generating all the necessary template dmb's, you may run ARFS2Batch. You will first be prompted to select your template dmb's, select all of them at the same time. ARFS2Batch will then quickly determine the FOV of all videos in the directory (if the dmb is in a different folder than the videos, you will be asked to find the videos). A small message box will appear to give you a summary of the FOV's. You will then be prompted to select all the videos that belong in the same DeMotion set as a given template dmb. This selection is limited by modality and FOV, so don't be concerned when you don't see your secondary sequences. Ctrl and Shift (what is it for Mac? Cmd?) shortcuts work during this selection. You will then be asked to set the ARFS parameters for each set (Fig. 4).

The minimum number of frames allowed per cluster should be set to balance eye motion and image quality. If you have high nystagmus and good image quality, you might set this really low, e.g., ~3. If you have low nystagmus but poor image quality, you should set this higher, 10-20. If, god forbid, you have high nystagmus and poor image quality, I recommend dropping this to 1 or skipping motion tracking all together (below).

Humans routinely select multiple reference frames to either, register slightly different parts of the retina, or to boost their chances of having a successful registration. In theory, the number of reference frames to try per video should work if set to 1, but occasionally, ARFS has a blunder, and it would be safer to select 2-3 frames. If all 3 fail, double check that the reference frames it's selecting look appropriate, if not, let me know. Otherwise, changing your registration parameters either manually or automatically, should help.

You should process the best frames overall in cases of really heavy nystagmus, otherwise, it should always work better to process the best frames from each cluster and sort it out later.

Motion tracking can be skipped in cases of heavy nystagmus. This will drastically reduce processing time, but will not provide any useful spatial information. In this case, only the first pass of motion tracking is conducted, as NCC's are still useful as an inference of distortion.
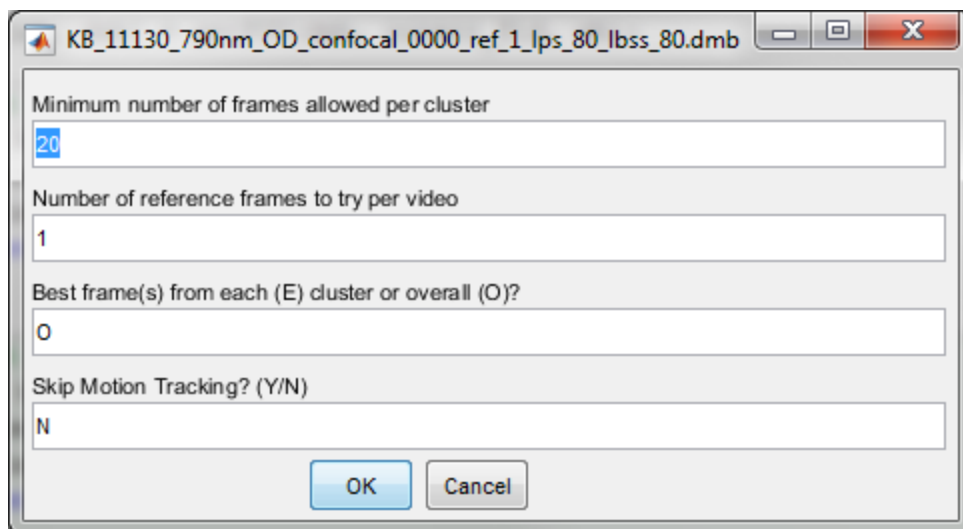

Fig. 4 ARFS parameter selection. Adjust cluster size threshold, number of frames to output, whether to generate a dmb from the least-distorted frame in each cluster or overall, and whether to skip motion tracking.

These parameters must be set for each template dmb used, and the lower two fields are case-insensitive.

Section 2.2: Outputs and Errors

In the event that ARFS2Batch experiences an error, it will be recorded in a text file with the session date and time followed by stdout.txt (short for standard output). It's worth quickly checking this text file for any errors reported before moving on to strip-registration. Additionally, it will output a file with the session date and time followed by arfsdata.mat. If any errors arise, it is very useful to send both of these files in an email for troubleshooting.

Most importantly, ARFS2Batch outputs dmb's with updated video names, desinusoid file names, and reference frames. These must then be dragged into DeMotion for strip-registration.

References

Arthur, D., & Vassilvitskii, S. (2007). *k-means++: The advantages of careful seeding.* Paper presented at the Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms.

Dubra, A., & Harvey, Z. (2010). *Registration of 2D images from fast scanning ophthalmic instruments.* Paper presented at the International Workshop on Biomedical Image Registration.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics, 20*, 53-65.

Yellott, J. I., Jr. (1982). Spectral analysis of spatial sampling by photoreceptors: topological disorder prevents aliasing. *Vision Res, 22*(9), 1205-1210.