

Instrument Science Report ACS 2014-03

Instrument Science Report WFC/UVIS 2014-23

# **hst2galign: an Automated Galaxy-based Alignment Routine**

---

Jay Anderson and Sara Ogaz  
October 10, 2014

---

## **ABSTRACT**

This brief document describes the procedure by which the individual images in the Frontier Fields program have been aligned in order to enable the self-calibration procedure (which will be described in a separate document). Along with this document, we will release the FORTRAN source code (`hst2galign`) that accomplishes the alignment. The source code is provided as-is, with no guarantee that it will work on any particular data set. However, it should work “out of the box” on datasets that are similar to the Frontier Fields.

## **1. INTRODUCTION**

The idea of the limited Self-Calibration procedure adopted for the Frontier Fields (FF) reductions of the Advanced Camera for Surveys (ACS) data is that each individual exposure has two main contributions: the photons that come down from the sky and the photons that are introduced by the detector (such as dark current and other related artifacts). The goal of Self-Cal is then to construct these two images: an image of the astronomical scene and an “image” that represents the contribution of the detector to each exposure. By doing this, we can remove the detector signature from the individual exposures.

Each FF dataset consists of a minimum of 20 and a maximum of about 80 exposures of the scene for each filter, with small dithers between the pointings. This constitutes the ideal set-up for a Self-Cal solution, since a given part of the scene always lands in many different pixels, allowing a very effective separation of the sky and detector contributions.

The Self-Calibration procedure relies on having a reliable forward and reverse mapping from each pixel in each exposure to the sky. This is what allows us to decompose each pixel into the sky and detector

contributions. There are very few stars in these images, so these transformations must be constrained by means of the galaxies. We describe in this document the basics of how the routine works. The program is not designed to work for every situation. In particular, it has not been stress-tested to see how few or how many images it will work on. Currently, it works only on images that are taken with the same detector through the same filter at nearly the same orientation. In principle, it could be made to work on images with different orientations and possibly even through different filters. Any such improvements will be left for future implementations.

## 2. TRANSFORMATION BASICS

The current version of the `hst2galign` routine described in this document operates outside of the `Astrodrizzle` environment. As such it uses different distortion solutions and different methods for measuring positions and constraining transformations.

The distortion solution that it uses is in the form of fits images: one image for  $x$  and one for  $y$ . Each image is the size of the detector (for ACS/WFC and WFC3/UVIS 4096×4096 pixels, and for WFC3/IR 1014×1014) and the pixel value in the image reports where the center of each pixel maps to in the adopted distortion-corrected frame. Using the two images in concert, we can map each pixel directly into a distortion-corrected frame.

Distortion-free frames can be shifted, rotated and stretched in so-called “conformal” transformations, and remain distortion-free. This gives us a lot of freedom in choosing a distortion-corrected frame. For the purposes of differential astrometry, it is easiest to work with a frame that is as close as possible to the detector itself, so Anderson & King (2006) for ACS and Bellini, Anderson, and Bedin (2009) for WFC3/UVIS chose an offset that maps the central pixel in the bottom chip (at [2048,1024] in ACS/WFC and WFC3/UVIS, simply [507,507] in WFC3/IR) to have the same coordinate in the distortion-corrected frame. We also constrained the scale and orientation such that the  $y$ -axis in the distortion-corrected frame is up and has the same pixel-scale as the detector at that location. The distortion correction, then, tells us where every other pixel maps to with respect to the  $y$ -axis in this central pixel.

There are three main coordinate systems that `hst2galign` works with. The user of the routine will not need to understand all of these details, but for completeness it is worth describing a bit about how the mapping is done from the individual exposures into the reference frame and back. The three main coordinate systems are: (1)  $\mathbf{x}_r, \mathbf{y}_r$  — the raw coordinates in each detector frame (for ACS and UVIS, we abut the two chips into a single 4096×4096-pixel system for convenience); (2)  $\mathbf{x}_c, \mathbf{y}_c$  — the distortion-corrected coordinates, and (3)  $\mathbf{u}, \mathbf{v}$  — the master-frame coordinates. We adopt the distortion-corrected frame of the first image (sometimes offset to ensure that the entire field of interest have positive coordinates) as the reference frame. So, in essence, we simply need to map each exposure into the frame of the first exposure. It is worth noting that these distortion-corrected frames are not identical to the frames accessed by the IDCTAB files used by `Astrodrizzle`, which needs a frame that is rigidly tied to the telescope’s V2/V3 system for the purpose of absolute calibration and allows dead-reckoning mapping. It is more important that the frame we use be easy to access and as close as possible to the detector frame.

We have already described the mapping from  $(x_r, y_r)$  to  $(x_c, y_c)$  — this mapping (and its inverse) are accomplished by means of the distortion-correction reference images<sup>1</sup>. The mapping from  $(x_c, y_c)$  to the master-frame coordinates  $(u, v)$  is accomplished by means of a simple 6-parameter linear transformation:

$$\begin{aligned} \mathbf{u} &= \mathbf{A} \cdot (\mathbf{x}_c - \mathbf{x}_0) + \mathbf{B} \cdot (\mathbf{y}_c - \mathbf{y}_0) + U_0 \\ \mathbf{v} &= \mathbf{C} \cdot (\mathbf{x}_c - \mathbf{x}_0) + \mathbf{D} \cdot (\mathbf{y}_c - \mathbf{y}_0) + V_0 \end{aligned}$$

Formally, it looks like there are eight parameters: four linear parameters (A, B, C, and D) and four offset parameters ( $x_0$ ,  $y_0$ ,  $u_0$ , and  $v_0$ ). However, it turns out that two of the offset-related parameters are arbitrary, in that they just tell us where an arbitrary but particular point — say,  $(u_0, v_0)$  — maps to in the other frame. These linear transformations are typically constrained by a set of what we call “point associations” wherein we have identified the same star (or galaxy) in two different frames and have a pair of coordinates that, within some error, correspond to the same point in space. In other words, we have N pairs of coordinates  $\{ x_n, y_n; u_n, v_n \}$ , and they can be used to compute a linear transformation between the two frames. The minimum number of stars (or galaxies) we can have is three, since that gives us three constraints for the three unknowns in each coordinate.

It is worth noting that these point associations are never perfect. Typically,  $(x_n, y_n)$  comes from a raw measurement in an individual exposure. It can have uncertainties related to: Poisson errors, errors in the distortion solution, and other sources of error in the values of the pixels that were fitted, as well as errors in the PSF model used to fit for the position. For bright stars with  $S/N > 100$  these errors total about 0.01 pixel and for fainter stars the errors tend to scale as  $(S/N)^{-1}$ . The reference-frame position  $(u_n, v_n)$  typically has smaller errors than the individually measured positions (since they tend to be averages of such positions), but they can have errors as well, and are rarely more reliable than 0.0025 pixel.

To determine the linear transformation that best describes the set of point associations  $\{ x_n, y_n; u_n, v_n \}$ , we use least squares to determine the optimal parameters in the above two coupled equations. To solve for the position parameters, we typically just set  $x_0$  to the average of the  $x_n$ 's, and do the same for  $y_0$ ,  $u_0$ , and  $v_0$ . In this way, we define the “arbitrary” point to be the centroid in one system, and the math makes it such that the other point is simply the centroid in the other system. We then solve for A, B, C, and D using standard least squares (see the subroutine `glob_fit6uuw` in the referred-to software). It is worth noting that this routine allows for the fact that all the point associations may not be of equal quality. We could very easily weight them according to either the signal-to-noise or the quality of the measurement. In the case of using galaxy positions for the point associations, it is important to note that some galaxies are fuzzier than others and they may not give high-accuracy positions. We will revisit this in Section 9.

There is one last general point worth mentioning. The linear transformations described above are general 6-parameter transformations. They account for more than a typical “conformal” transformation, which includes terms for x- and y-position offset, rotation, and a scale change<sup>2</sup>. We have seen above that such “conformal” transformations do not introduce distortion. The transformations used here *can* introduce distortion, since they have two additional linear terms, often called “skew” terms. One of these terms allows for scale difference between the  $x$  and  $y$  axes and the other allows for a non-perpendicularity

<sup>1</sup> Many of these images can be accessed at the following URL: <http://www.stsci.edu/~jayander/GCLIB>.

<sup>2</sup> In terms of the parameters described above, conformal transformations have  $A=D$  and  $B=-C$ .

between the axes. The current version of `Astrodrizzle` operates under the assumption that all the distortion terms are accounted for in the distortion solution itself, so it does not allow the flexibility of a full 6-parameter linear transformation. The solution used here is flexible enough to absorb some errors or variation in the linear part of the distortion solution.

We note, however, that in data sets such as we are analyzing here, where all the observations are taken at the same orientation and presumably a similar epoch, then there is almost no differential skew among the images, so this package and `Astrodrizzle` can achieve similar results as far as this issue is concerned. It is when comparing data sets taken with different instruments, at different orientations, or at different epochs (the distortion solution for ACS changes over time) that unaccounted-for skew can enter into the transformations and cause problems with routines that are limited to conformal transformations.

### 3. INITIAL MEASUREMENT OF THE SOURCES

The first stage of the `hst2gal` procedure is to find and measure the sources in individual exposures. There are three parameters that regulate the source-finding algorithm: `P`MAX, `F`MIN, and `H`MIN. We go through each pixel in each exposure and identify potential sources in those pixels that: (1) are lower than `P`MAX, which is usually set to the level of saturation, (2) have more than `F`MIN counts over the sky-background level within their inner  $3\times 3$  pixels, and (3) have no brighter pixels within a radius of `H`MIN pixels. We then determine a simple centroid position for the object and the flux-over-sky within the central  $5\times 5$  pixels and add it to the list. If an optional PSF is provided, then the source will be fit to the PSF and may be accepted or rejected based on the quality of its PSF fit. Finally, since we could generate an extremely long list this way, particularly for crowded fields, we retain only the brightest `N`STMAX stars. This list is output in the `.xym` file for each exposure. This file contains the raw and distortion-corrected positions, fluxes in terms of instrumental magnitudes, the results of the PSF-fits (if a PSF has been provided), and other useful information.

At this point, we assume that all observations were taken at the same orientation, or that a WCS has been used that orients and scales everything into a common system, such that the frames of the exposures are related largely by simple offsets. We adopt the distortion-corrected source list from the first exposure as the reference source list and use a hash-table approach to find the offset between the individual exposures and this list. This allows us to do a quick cross identification between the reference list and the source list for each exposure, thus giving us a list of point-associations with which we can determine the linear transformation from each exposure into the reference frame. This transformation is good, but not perfect, in that it is limited by the crude centroid-nature of the input source positions and the fact that the reference positions come from the first exposure and suffer from these errors as well.

### 4. A LIST OF COMMON SOURCES

The initial reference list came directly from the first exposure. There could be sources of interest that were not found in the first exposure, either because they landed on artifacts or off the image entirely. We go through another wave of finding wherein we use a “peak map” (see Anderson et al. 2008) to identify common sources — sources that are found in a minimum number of exposures (`N`PKMIN). This provides a list of sources that is not in any way biased by the first exposure. Since some of these newly

found sources might map to negative coordinates in the initial exposure-#1-based reference frame, we shift this new frame to ensure that all the sources will have positive coordinates.

## 5. USING ITERATION TO REFINE THE TRANSFORMATIONS

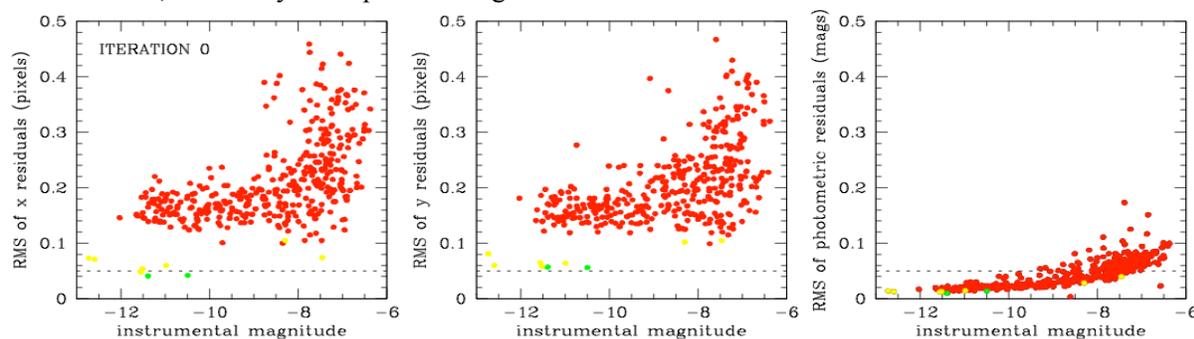
Now that we have an improved list of sources with positions in the reference frame, we cross-identify them with the stars found in the  $M$  individual exposures (from Section 3). We then use iteration in a virtuous cycle to improve the positions and transformations.

The first stage of the iteration involves determining an average position for each source in the reference frame. We use the currently best transformation coefficients for each exposure ( $m$ ) to take the distortion-corrected position for each source ( $n$ ) and determine an estimate for its reference-frame position:  $(\mathbf{u}_{nm}, \mathbf{v}_{nm})$ . For each source, we take these  $M$  estimates and find a robust average position, along with a robust average flux. By “robust” we mean that we find an average, determine an RMS, and iteratively reject the outermost point until all the points agree to within  $\sim 4$ -sigma or better. As observations are rejected, they are flagged so that they will not be used in the transformations.

The RMS of the position residuals is used to tell us how accurately each source can be measured in each exposure. We give all sources a weight of 1.00 if they are measured to better than 0.05 pixel, but above this the weight goes down quadratically as  $0.05 \times RMS^{-2}$ .

These improved average reference-frame positions are then used to improve the transformations, this time using the weights and flags determined from the above procedure. After about five iterations of alternatively determining the average positions and then re-computing the transformations, the routine converges on a set of average positions and transformations that can be trusted to relate the pixels of each exposure to the sky.

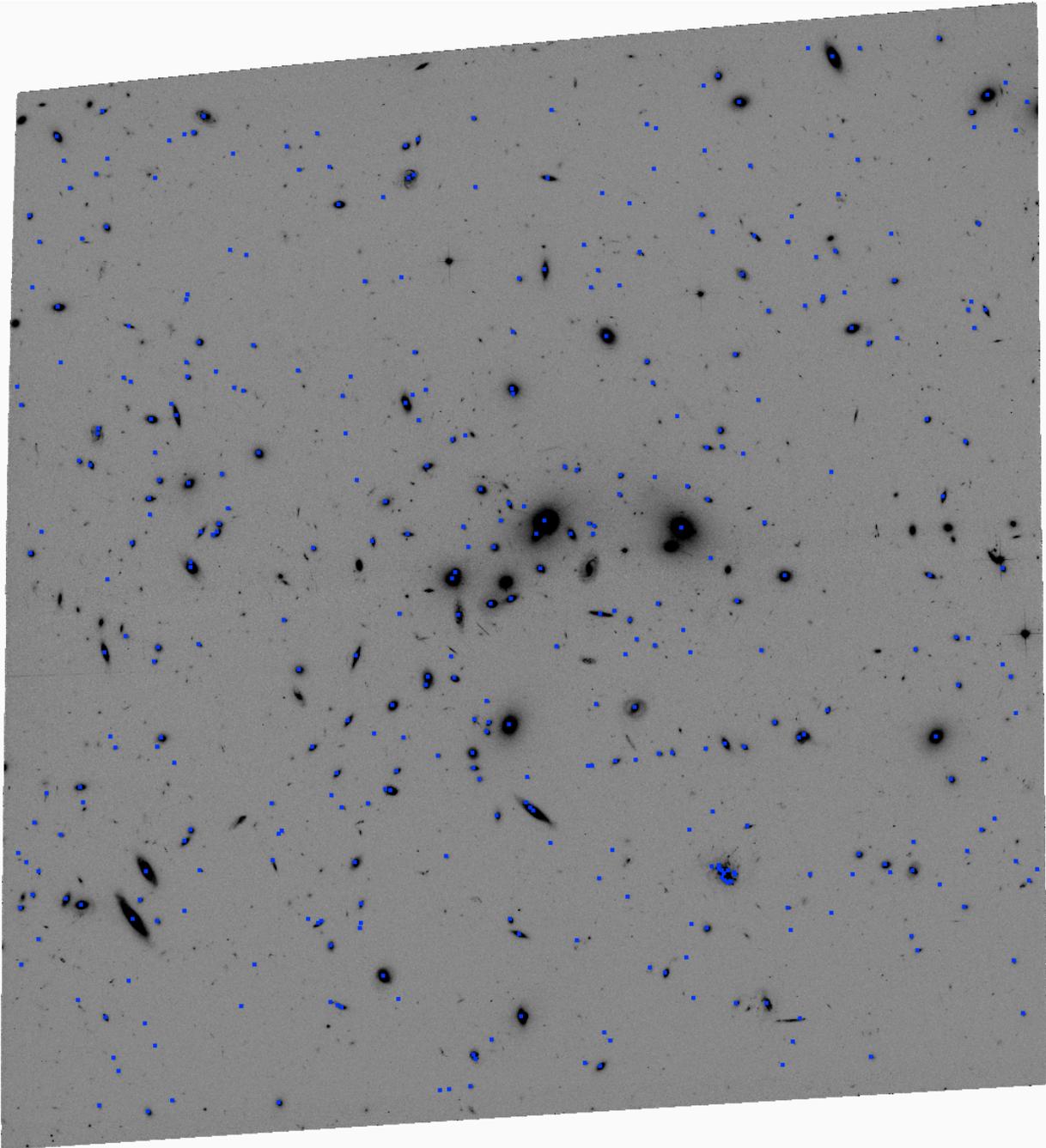
**Figure 1** shows the position and photometric residuals as a function of source brightness. It is clear that the centroid-based position measurements are good to about 0.05 pixel for the brightest galaxies and get worse from there. The photometry is good to about 0.02 magnitude (about 2%). Objects with position residuals better than 0.05 pixel are assigned a weight of one, with the weight decreasing quadratically as  $w \sim (RMS/0.05)^{-2}$  above that. Green points have weights of 0.90 or better and red points have weights of 0.25 or worse, with the yellow points being somewhere in between.



**Figure 1:** This RMS position and photometric residuals after the collation of the centroid-based positions. Each point represents one of the  $\sim 400$  sources found in the field. The x and y residuals are shown in the left two panels, and the photometric residuals in the right panel. The predominance of the red colors mean that most objects had large measurement errors ( $>0.1$  pixel). Although the centroid-measured positions have a large scatter from image to image, the simple aperture photometry yields a surprisingly consistent measurement from exposure to exposure.

## 6. GENERATION OF A SANITY-CHECK IMAGE

Although it is not the focus of the routine `hst2gal` to generate a stacked image, it does generate a sanity-check stack in the adopted reference frame. **Figure 2** below shows the stack for the F814W Frontier Field observations of the Abell 2744 cluster. The 396 “common” sources are indicated in blue.



**Figure 2:** The representative stack made of the Abell 2744 cluster field. The stack is not photometrically rigorous, but it does help us make sure the routine is finding good galaxies.

## 7. “GSF” TEMPLATES

If we had started with PSF-measured positions of only stars, then there would be no way to improve the raw positions measured for each star, and the above procedure would provide the best estimate of the average reference-frame positions and transformations. The position residuals could tell us how good the PSF model is, but the routine is not set up to extract PSFs.

Typical extragalactic fields (such as the Frontier Fields) do not contain enough stars to allow us to base our transformations on the stars. Furthermore, stars are known to move, so they do not provide stable points of reference for data sets taken over extended epochs. Therefore, we often must rely on galaxies. In the initial finding stage, since we knew nothing about each galaxy’s size and shape, we used crude centroids to estimate their positions. As is true for stars, it is possible to get much better position measurements for galaxies if we have an accurate template for them and can do a least-squares-type fitting of the pixels in their profiles to the template.

The PSFs in the HST images of the three detectors being considered here are all undersampled. This means that special care must be made in constructing models to describe how the light of point sources is distributed. A single exposure cannot tell us how concentrated the PSF is (see Figure 1 in Anderson & King 2000), so we must combine multiple dithered exposures if we hope to construct an accurate PSF model. Since, all stars are the same, one can construct a single PSF model based on the images of a great many stars.

Using galaxies for astrometry is quite different. In some ways it is easier, in other ways it is more challenging. Galaxies can be *easier* to model in that they are resolved, and as such they do not have the high-frequency structure that is present in the PSFs. We can therefore use more coarsely sampled models for the galaxies. Of course, to the extent that galaxies have point sources at their centers, they may have the same structure present in the PSF, but it turns out that not many galaxies have dominant point sources at their centers at HST resolution. The thing that makes galaxies *harder* to model is that every galaxy is different. With stars, we can use 100 to 1000 stars in (say) a  $500 \times 500$ -pixel region of an image and combine all their profiles together from all the contributing images into a single solution, since they should all have the same PSF. This cannot be done for galaxies. We have only as many estimates of a given galaxy’s GSF (galaxy-spread function) as we have contributing exposures.

It is worth mentioning here that when we are discussing “astrometric” sources, we are not necessarily talking about distinct galaxies. We could be talking about parts of large galaxies, such as its globular clusters, knots in its spiral arms, etc. — anything that is sharp enough to provide a good astrometric handle can lend itself to a GSF model.

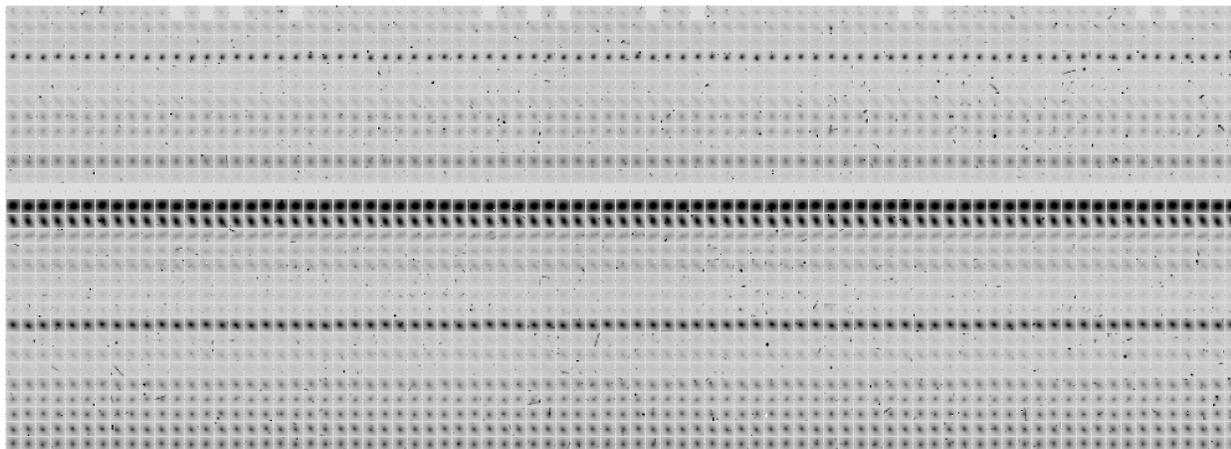
The concept of a “handle” is an important one (see discussion in Mahmud & Anderson 2008). When we solve for a PSF, we can simply adopt its center as its “handle”. When we say a star is at a particular coordinate, we simply mean that the center of the best-fit PSF model maps to that location. We could define the center to be the brightest point in the PSF, but that is ill-defined since the spatial gradient is by definition flat at the peak. We could use the technical centroid (the first spatial moment of the light distribution), but that can be pulled off of the obvious center by asymmetric light in the outskirts, so such definitions naturally require some finite window. We could use some point of radial symmetry, or could do a Gaussian fit and take the center of the Gaussian to be the center of the PSF. In the end, how we

define the center is entirely arbitrary — even for stars. But that does not mean that it is not important. When dealing with spatially varying PSFs in particular, it becomes quite important to adopt the same centering convention for all the PSFs so that we can interpolate between them without introducing blurring.

Adopting a “handle” for galaxies is not as important as it is for stars, since the galaxy profile should not vary with position on the detector (like a PSF does). The handle is simply a point within the GSF that we will call the galaxy’s center. When we solve for the position of the galaxy, we are saying that that point in the galaxy’s profile is located at coordinate  $(x, y)$  in the image. We model the GSF on a  $\times 4$ -supersampled  $41 \times 41$ -point grid, analogous to the PSF grid, and adopt the center of the grid as the galaxy’s “handle”. Even though the location of the handle is arbitrary, it is nevertheless useful to have it correspond to something physical, since we may want to compare positions for the galaxy measured in different filters and we want those positions to be as similar as possible<sup>3</sup>. We generally adopt the point of radial symmetry within a radius of 2.5 pixels as the center. When we solve for a galaxy profile, if its “center” is not at the center of the GSF grid, we resample the grid such that the center of the profile will be at the center of the resampled grid.

## 7. THE NITTY-GRITTY OF TEMPLATE-CONSTRUCTION

Now that we have laid the theoretical groundwork for construction of the galaxy templates, we can begin to measure them. To construct the templates, we extract an  $11 \times 11$ -pixel raster about each object in each exposure. We also use the transformations to determine where the center of each pixel in each individual-image raster maps to in the reference frame. [Figure 3](#) shows the array of extracted rasters; each row corresponds to the same galaxy, and each column to the same exposure. Cosmic rays and other image defects are clearly present in some rasters, but they rarely impact the same galaxy image at the same location.



**Figure 3:** Extracted rasters from for 30 galaxies for each of the ~80 exposures.

---

<sup>3</sup> Of course galaxies have different profiles in different filters, but to the extent that they are similar, it makes sense to desire a consistent handle.

The GSF template itself is made up of a grid of  $41 \times 41$  points that is super-sampled by a factor of 4 relative to the reference-frame pixels. To determine the value of each gridpoint, we start by going through the exposures one by one and finding the pixel in each raster that lands closest to the mapped location of that gridpoint. We take that pixel value as the initial estimate for the gridpoint value. We determine a robust average of these estimates from the  $M$  contributing images and assign that average value to the gridpoint. We do this for the entire array of gridpoints. We then smooth the grid with a 2-D kernel that respects quadratic variations in the inner parts (within  $\sim 4$  pixels) and becomes planar in the outskirts. This ensures that we can minimize noise in the grid while respecting the structure that we believe might be real.

We then improve the gridpoint values by iteration. We remove from each observed pixel in the  $11 \times 11$  rasters the GSF model at its location (determined by using a bi-cubic interpolation of the gridpoints) and then construct a robust average of the residuals in the same manner as above when we used the pixel values themselves. This iterative procedure is similar to *iDrizzle* (Fruchter 2011), and it enables us to construct a set of fiducial gridpoint values, even though we do not in general have samples that map to the exact gridpoint locations. At the end of each iteration, we also determine a point of radial symmetry in each GSF and resample the grid such that the point of symmetry is at the center of the grid.

We perform ten iterations to converge upon the GSF model for each source. Things generally converge within one or two iterations, but the iterations run fast enough on modern computers that ten is not onerous.

## 8. FITTING IMAGES OF GALAXIES WITH GSFs

It is worth noting that the above GSFs are constructed in the distortion-corrected reference frame. As such, they are somewhat different from library PSFs, which are constructed in the `flt` frame. The GSFs are, though, *used* in the `flt` frames to determine the location of the “handle” in each detector-based frame.

To measure a source in an exposure, we extract an  $11 \times 11$  raster about the central pixel of the object from the `flt` image. For each pixel, we use the current-best transformations to determine the master-frame-mapped position for that pixel, so we have an association of  $(i, j ; u, v)$  for each of the 121 pixels. We then determine a local linear mapping from the `flt` frame to the reference frame. We use the linear terms of this mapping to convert the GSF from the distortion-corrected master frame into the distorted `flt` frame.

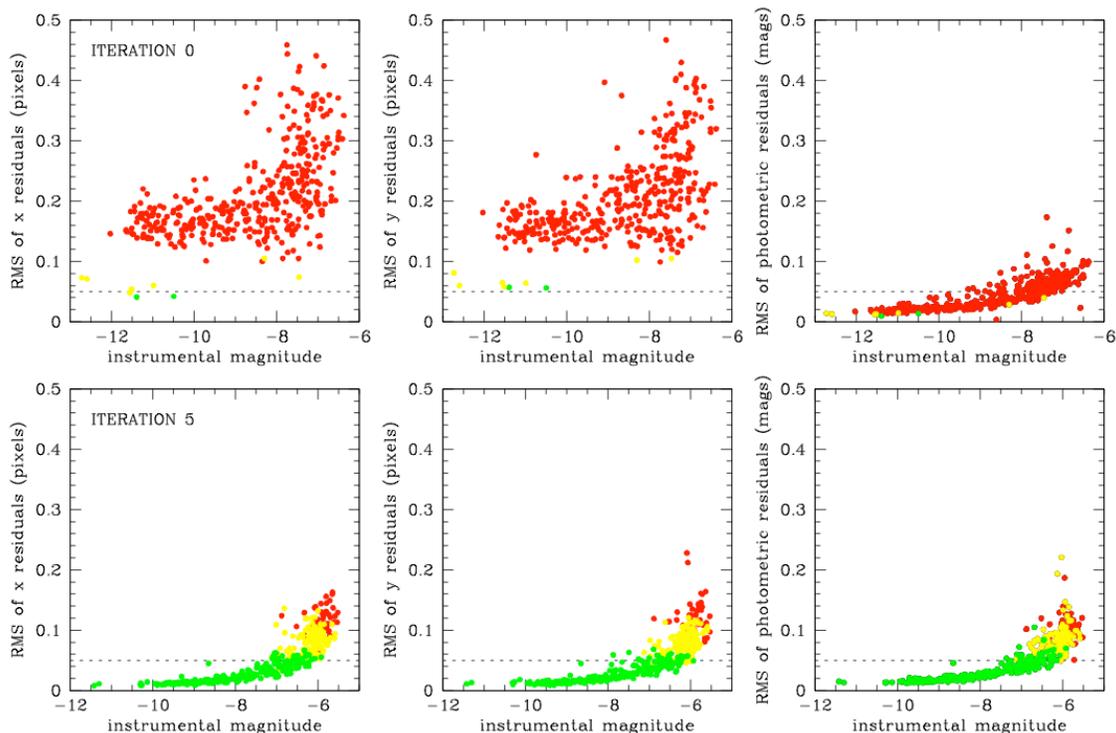
We then fit this converted GSF via least squares to the  $5 \times 5$  central pixels of the object for  $(\mathbf{x}_r, \mathbf{y}_r, \mathbf{z}, \mathbf{s})$ , where  $\mathbf{z}$  is the flux of the object and  $\mathbf{s}$  is the sky. This fit is non-linear in  $\mathbf{x}_r$  and  $\mathbf{y}_r$ , but at a given  $\mathbf{x}_r$  and  $\mathbf{y}_r$ , it *is* linear in  $\mathbf{z}$  and  $\mathbf{s}$ . Rather than do a complicated gradient search, we simply evaluate the quality of fit at a grid of  $\mathbf{x}_r$  and  $\mathbf{y}_r$  values spanning  $\pm 1$  pixel in each coordinate sampled by 0.01 pixel. At each trial location, we solve for  $\mathbf{z}$  and  $\mathbf{s}$  then evaluate the quality of the fit. We adopt as the best-fit position, flux, and sky, the solved-for values at the location where the linear-correlation coefficient between the model and data is closest to one. The result of this fitting is an  $(\mathbf{x}_r, \mathbf{y}_r)$  position in the raw `flt` frame that corresponds to the center of the GSF template, the “handle” for the galaxy. This position can then be corrected for distortion and used to improve the transformations.

## 9. THE GRAND ITERATIONS

In order to construct accurate templates, we need access to good transformations. However, we need good templates in order to measure good positions and thus improve the transformations. This suggests a virtuous-cycle iterative approach.

We use the initial centroid-based positions to construct the internal transformations described in Section 4; this also gives us an average position for each source in the reference frame. We then use these average positions and the transformations to construct a template for each object. This template is in turn used to determine an improved position for each object in each exposure, and these positions are used in the iterative scheme of Section 5 to determine a new average reference-frame position for each object and thus improve the transformations. This outer iterative procedure is repeated five times.

**Figure 4** below demonstrates the improvement in our position measurements. The top panels show the same points for the initial iteration (shown in **Figure 1**). The bottom panels show the results after the fifth iteration. Here, there are many more green and yellow points than before, so we have a great many objects with extremely good positions. In these deep F814W ACS/WFC exposures of a typical galaxy field, we have 50 or more sources with a measurement quality better than 0.02 pixel, and 200 or so sources with positions better than 0.05 pixel (the threshold between green and yellow). The large number of objects with high-precision position measurements means that our transformations should be good to better than 0.01 pixel, at least as far as random errors are concerned.



**Figure 4:** The improvement in position measurements from centroid (top panels) to GSF-fit positions (bottom panels).

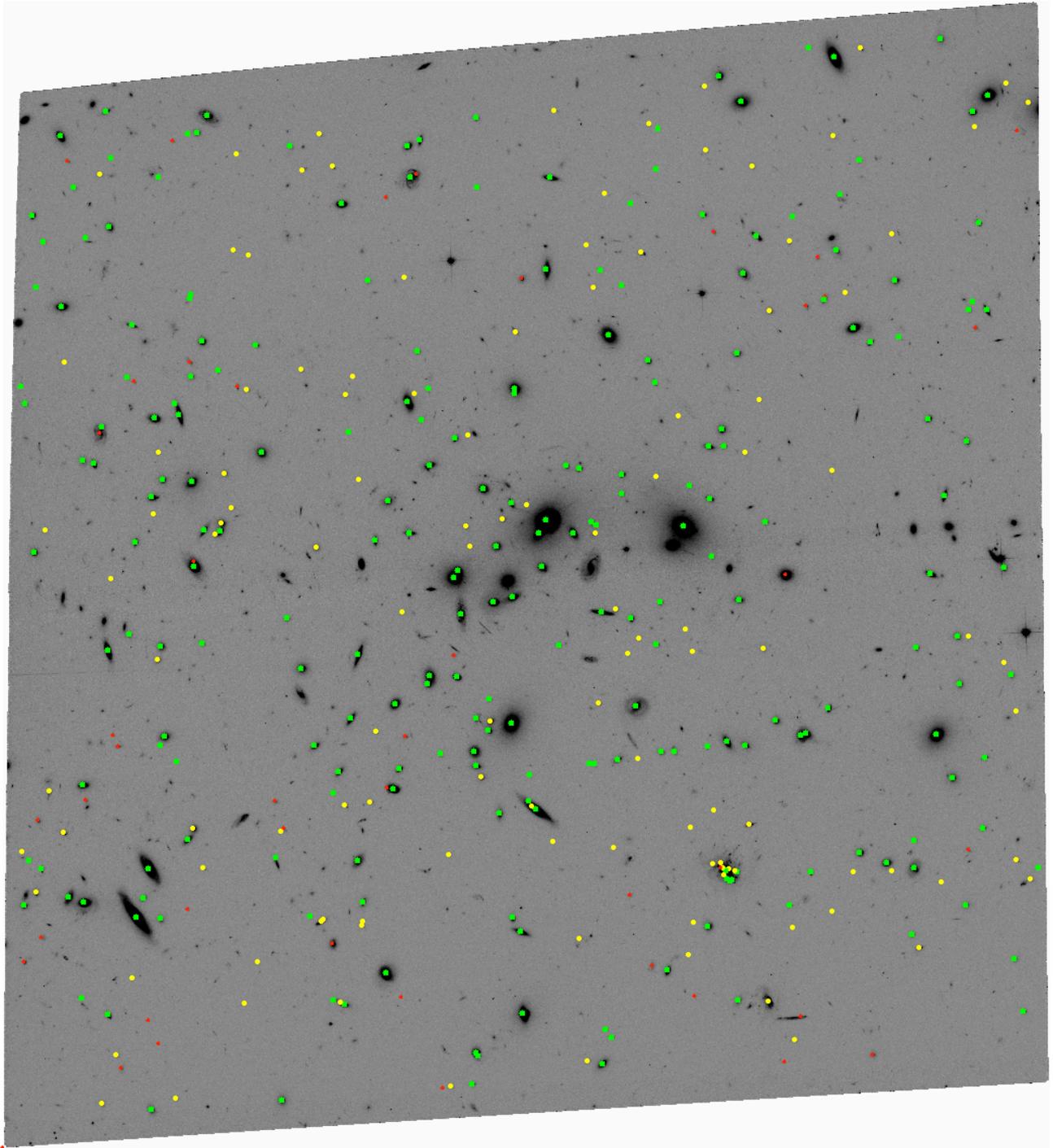


Figure 5: This shows the same image as in Figure 2, but with the galaxies color-coded by how good they are for measuring positions. As in Figure 4, green-coded galaxies are the best, with position residuals better than about 0.05 pixel and weights greater than 0.90. The red-coded galaxies are the worst with residuals greater than 0.10 pixel and weights less than 0.25. The yellow galaxies are of intermediate quality.

Figure 5 shows the whole field, with the galaxies now identified by quality. Most of the galaxies are green, indicating that they can be used to get positions to better than 0.05-pixel accuracy.

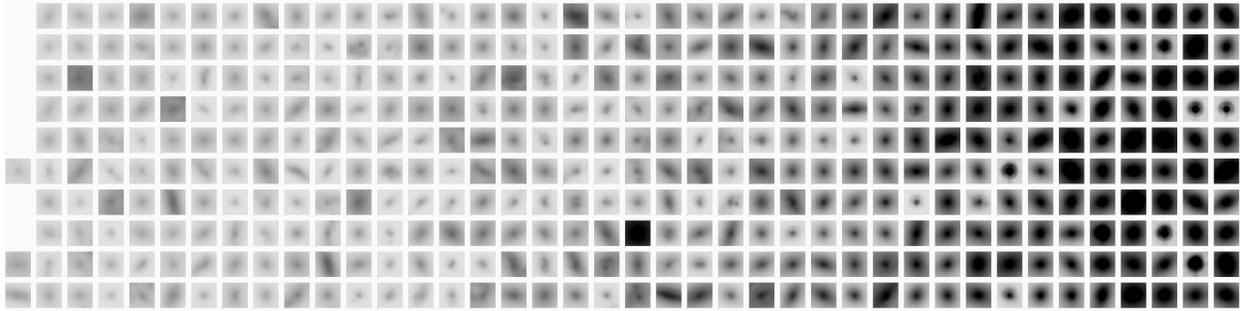


Figure 6: This shows all the galaxies used, ordered by the astrometric measurement quality, with the best galaxies being on the right.

Figure 6 shows the array of sources used. As expected, the brighter the source, the better it can be measured. There are a few bright objects that have broad profiles in one or both dimensions. These galaxies have lower astrometric quality. There are a few stars here, but most objects are nicely resolved galaxies. There is an option for the routine to use only stars or to select against using stars, if a PSF is provided to the routine.

## 10. THE OUTPUT

The `hst2galalign` software can be found at the following URL:

[http://www.stsci.edu/~jayander/FF\\_SOFTWARE/hst2galalign](http://www.stsci.edu/~jayander/FF_SOFTWARE/hst2galalign)

It is designed to be compiled with GFORTRAN. Here is an example of the command-line run of the routine (note that there can be no spaces after the line-break character “\”):

```
./hst2galalign.2014.12.09.e \  
HMIN=3          \  
FMIN=25         \  
PMAx=99999999 \  
CRLIM=0.0      \  
INST=WFC3IR    \  
"FILEGCX=GCLIB/WFC3IR/wfc3ir_gcx.fits" \  
"FILEGCY=GCLIB/WFC3IR/wfc3ir_gcy.fits" \  
NPKMIN=0.50    \  
MKSTACK+      \  
DOWCS+        \  
EXPT_0=1000.0 \  
../00.data/*_flt.fits
```

It was written for a specific purpose: to provide a framework for the self-calibration routine to inter-relate the exposures in the data set. The goal of the routine, then, is to generate a “mat” file for each exposure. This “mat” file consists of a list of sources that can serve as “point-associations”. For each source we

have a position in the particular exposure (measured from the galaxy templates) that we can associate with a position in the reference frame (the average position), along with a weight that tells us how good this point-association is. These point associations implicitly define the transformation from each exposure into the reference frame. The self-calibration routine will read in the point associations from the `mat` files to constrain the transformations from each image into the reference frame.

The main diagnostic output file is named `LOGC.xym2bar.UVZEEE.05`. It reports the average positions and fluxes, as well as the RMSs of the individual observations for the fifth (and final) iteration. This is a good file to look at to make sure that the objects were identified and well characterized. [Figure 4](#) shows the output of this file. After a successful run, there should be a trend where there are a few bright objects that are relatively well measured and a larger population of faint objects that are less well measured. These points can be plotted on the output stack image (`LOGA.stack_bar.fits`) using the DS9 region file `LOGH.MAST2REG.reg`. Other auxiliary files show the array of galaxies (such as `LOGC.show_galarr05.fits` as is shown in [Figure 3](#), or `LOGC.rast2tem_srt.05.fits` as in [Figure 6](#)) in various stages of iteration and adjustment.

## 11. THE SOFTWARE: WHAT IT DOES AND WHAT IT DOESN'T DO

As mentioned above, the software routine `hst2galign` is written in FORTRAN. It was written to solve for the image-to-image alignment for Frontier-Fields-like data sets where there are between 20 and 80 exposures taken through a given filter with a given detector at nearly the same orientation at nearly the same epoch. It analyzes the dataset for only one filter at one time, and independently for each filter. It identifies the common sources that are bright enough to provide the astrometric registration and then performs a virtuous iterative process of measuring templates for each galaxy and using those templates to improve the alignment of each exposure to the adopted reference frame.

One of the novel contributions of this routine is the fact that it provides a clear demonstration of the astrometric quality of the transformations. We have a number of sources and are able to characterize each one in terms of the quality of its astrometric handle. This allows us to optimize the transformations like never before. We can determine whether the measurement of a particular source in a particular image is “good” in terms of whether its residual is consistent with the other measurements of the same source and different sources, thus allowing us to discard bad individual observations.

The transformations shown above indicate that there are enough bright, high-quality galaxies in wide-band half-orbit exposures to register images to better than 0.01 pixel, provided that the distortion solution and linear transformations are accurate enough to allow this. In fact, there are probably enough galaxies in 500s exposures to allow this precision. There are fewer high-quality galaxies available as the wavelengths get shorter. In the parallel Frontier-Fields observations, the half-orbit F435W exposures have just enough galaxies for 0.01-pixel precision. The bluer filters in WFC/UVIS (for instance the F275W observations of the UDF) have considerably fewer usable galaxies, so not all data sets will benefit from this approach. The WFC3/IR observations contain an abundance of astrometric galaxies, such that registering them to 0.01 WFC3/IR pixel should not be a challenge (see [Figure 7](#) below).

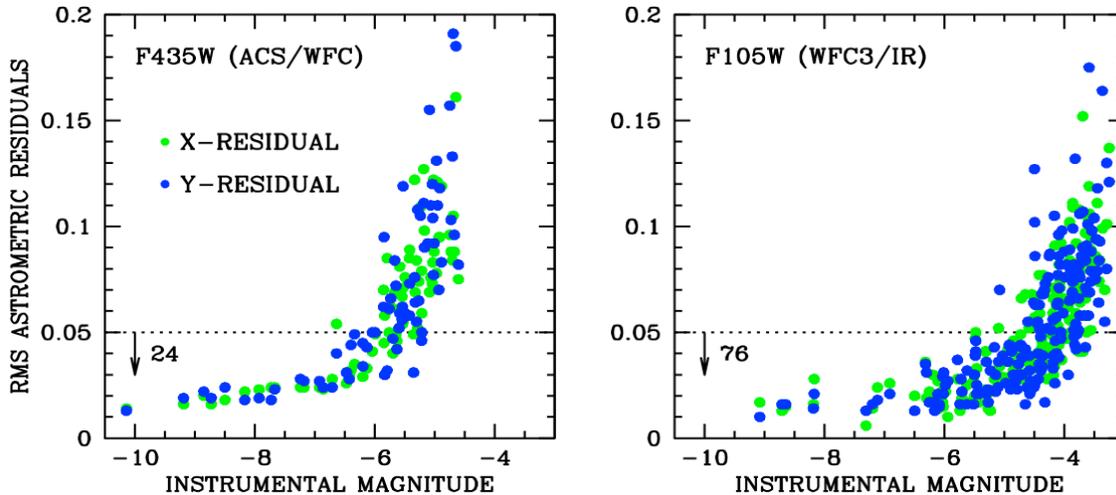


Figure 7: Astrometric residuals in each coordinate for the parallel field for MACS-01416 (essentially a empty field). There are 24 high-quality galaxies in the half-orbit ACS/WFC F435W exposures and 76 high-quality galaxies in the F105W WFC3/IR exposures.

While the routine’s achieved precision is good, there are many things that the routine does not do. It does not at this point generate the reference list in a frame that can be directly compared with drizzle products. The default reference frame that it adopts is based as closely as possible on the distortion-corrected frame of the first exposure, so it is not oriented in any particular way with respect to North. There is a “DOWCS+” option that uses the WCS information in the `flt`-image headers to construct the reference frame in a frame with North up and an attached WCS, but that does not necessarily match it up with a particular frame that `Astrodriizzle` has made.

When this routine is run on different data sets taken through different filters, it does not use a consistent reference frame, but it rather adopts a frame that is based on the first exposure of each data set. It would not be hard to take the reference-frame positions as given (rather than iteratively solving for them as is currently done), but that option has not yet been created. Such a procedure would allow the alignment to be done in the same frame for multiple filters. To the extent that the astrometric galaxies have the same point of symmetry in the different filters, such a procedure would provide excellent cross-alignment among the filters. It is worth noting that just because a frame has an attached WCS does not mean that it is perfectly aligned with other frames with a WCS. There can be unaccounted for shifts between filters and errors in the WCS related to imperfect guide-star positions. Absolute alignment is almost never as good as differential alignment.

The routine has not yet been configured to work with exposures taken at different epochs. This would require careful attention that no stars be used in the reference net, as we cannot count on them being motionless in an inertial frame<sup>4</sup>. We have modified the code to allow for stars to be used exclusively or

---

<sup>4</sup> Field halo stars typically move about 0.5 ACS pixel over 10 years.

not used at all, but there are several more modifications that will be required before we can use the routine to *measure* absolute motions of stars with respect to the background galaxies.

We have not explored how many images are needed to construct useful templates for astrometric registration. For the Frontier-Field project, it certainly works with 20 or more images (up to 80, but more should also be possible). It would be worth exploring whether the alignment works well for as few as four (or perhaps even two) exposures, but this parameter space has not yet been explored.

There is no direct route to using these positions in order to inform the TweakReg stages of `Astrodizzle`, and it is beyond the scope of this document to blaze that path. In principle, it should not be terribly difficult, since the routine provides a “mat” file for each exposure that gives a validated and consistent position for each reliable object in the raw exposure frame, and this is what TweakReg starts with. However, for TweakReg to realize the full accuracy of these measurements, particularly when comparing images taken at different orientation or in different epochs, it would need to allow full 6-parameter linear transformations, rather than the more limited 4-parameter “conformal” transformations that `Astrodizzle` is currently restricted to. It would also be good for TweakReg to allow for a variable weighting, since some galaxies clearly provide better astrometric constraints than others, and we can’t always afford to throw out all galaxies that aren’t in the top-tier.

## 12. POTENTIAL IMPROVEMENTS

The previous section went over the things that this routine does, and the things it does not do. Here, we consider how it might be modified to be even more versatile.

It would be relatively easy to modify the routine to generate a consistent list of reference-frame positions for observations taken through different filters and even with different instruments. This would require two changes. In the initial stage where the cross-matching is done between the reference list and the list of sources extracted from a given exposure, the search procedure would have to allow for changes in orientation and (perhaps) even scale. Then we would have to modify the code that solves iteratively for the average positions, transformations, and GSFs to take the reference positions as fixed and solve only for the transformations and GSFs.

One could imagine perhaps making these galaxy templates into some kind of elaborate source catalog, one that not only gives standard parameters of the sources but also puts them into a standard array of pixels. One could also include in this package for each source, the original pixels in the `_flt` images that contributed to the stack of the source along with the local mapping to the RA/Dec or reference frame, and perhaps even PSFs, so that users could develop their own analysis tools for restacking, point-source fitting, or deconvolution.

This routine could be designed to construct cross-filter and cross-instrument reference frames. This could be done simultaneously or in a piecemeal way. For instance, we could take the “best” filter (ie, the one that should have the best registration) and use it to generate a reference frame that the other filters could be matched up against. This might also allow the routine to deal with observations at different orientations or epochs, but that would require the reference frame to be WCS-calibrated, so that observations through different filters at different orientations could be cross-matched easily.

All these improvements are good ideas, but they are beyond the scope of the current `hst2galalign` code.

## REFERENCES

- Anderson, J., King, I. R., Richer, H. et al. 2008 AJ 135 2114 “Deep ACS Imaging in Globular Cluster NGC6397: Reduction Methods”
- Anderson, J. and King, I. R. 2000 PASP 112 1360 “Toward High-Precision Astrometry with WFPC2 1.: Deriving an Accurate Point-Spread Function”
- Anderson, J. and King, I. R. 2006 ACS/ISR 2006-01 “PSFs, Photometry, and Astrometry for the ACS/WFC Detector”
- Mahmud, N. and Anderson, J. 2008 PASP 120 907 “Using Resolved Galaxies in HST Images to Measure Absolute Proper Motions”
- Fruchter, A. S. 2011 PASP 123 497 “A New Method for Band-Limited Imaging with Undersampled Detectors”