The Astrophysical Journal, 790:87 (14pp), 2014 August 1 doi:10.1088/0004-637X/790/2/87

SpArcFiRe: SCALABLE AUTOMATED DETECTION OF SPIRAL GALAXY ARM SEGMENTS

Darren R. Davis and Wayne B. Hayes
University of California, Irvine, CA 92697-3435, USA; drdavis@uci.edu, whayes@uci.edu
Received 2014 March 8; accepted 2014 June 4; published 2014 July 8

ABSTRACT

Given an approximately centered image of a spiral galaxy, we describe an entirely automated method that finds, centers, and sizes the galaxy (possibly masking nearby stars and other objects if necessary in order to isolate the galaxy itself) and then automatically extracts structural information about the spiral arms. For each arm segment found, we list the pixels in that segment, allowing image analysis on a per-arm-segment basis. We also perform a least-squares fit of a logarithmic spiral arc to the pixels in that segment, giving per-arc parameters, such as the pitch angle, arm segment length, location, etc. The algorithm takes about one minute per galaxies, and can easily be scaled using parallelism. We have run it on all ∼644,000 Sloan objects that are larger than 40 pixels across and classified as "galaxies." We find a very good correlation between our quantitative description of a spiral structure and the qualitative description provided by Galaxy Zoo humans. Our objective, quantitative measures of structure demonstrate the difficulty in defining exactly what constitutes a spiral "arm," leading us to prefer the term "arm segment." We find that pitch angle often varies significantly segment-to-segment in a single spiral galaxy, making it difficult to define the pitch angle for a single galaxy. We demonstrate how our new database of arm segments can be queried to find galaxies satisfying specific quantitative visual criteria. For example, even though our code does not explicitly find rings, a good surrogate is to look for galaxies having one long, low-pitch-angle arm—which is how our code views ring galaxies. SpArcFiRe is available at http://sparcfire.ics.uci.edu.

Key words: galaxies: fundamental parameters – galaxies: spiral – galaxies: structure – methods: data analysis – methods: observational – techniques: image processing

Online-only material: color figures

When you can measure what you are speaking about and express it in numbers you know something about it; but when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely advanced to the stage of science.

Lord Kelvin (1824–1907)

1. INTRODUCTION

The Hubble Ultra Deep Field (HUDF) represents about 1/13,000,000 of the celestial sphere and contains about 10,000 galaxies, suggesting that the entire sky contains upward of 10^{11} galaxies at the resolution and depth of the HUDF. Gaining quantitative structural information for this number of galaxies will require automated methods.

Although in widespread use, existing classification systems such as the Hubble (1936) or de Vaucouleurs (1959) systems are still fundamentally subjective and qualitative. To enable quantitative comparison with theoretical models of galaxy structure, or of structure evolution on cosmological scales, objective quantitative systems are needed.

There exist a number of previous methods that partially automate the process of the quantitative structural description of galaxies. BUDDA (de Souza et al. 2004), GIM2D (Simard 1998), and GALFIT (Peng et al. 2002) can use brightness profiles (e.g., Sérsic 1963) to model the light curves of various components of the disk, bulge, and bar. A more recent version of GALFIT (Peng et al. 2010) can also be used to describe many more structures, including spiral arms. Although GALFIT can automatically use χ^2 fitting to find the best fit of a given model to an image, it still requires significant human effort to carefully specify the number and parameters of all the model components. A two-dimensional (2D) fast Fourier transform can be used to estimate the dominant pitch angle and the most probable number of major arms (Davis et al. 2012), but it assumes symmetry in the spiral arms and requires careful human supervision (Seigar & James 1998). Pitch angles have also been estimated manually across small samples (e.g., Ma 2001). Au (2006) and Perret et al. (2009) can each fit a symmetric, two-armed, barred spiral model to the image, but many galaxies do not conform to such a model. Ripley (1990) describes arms as chains of line segments, but the arms must be attached to a bar or core, and their initial positions must be specified manually. Ganalyzer (Shamir 2011) looks for spiral structure by finding intensity peaks in an angle versus radial distance plot. It is very fast, but considers two arms at most, and is oriented toward producing a continuous measure between spiral and elliptical classifications, as well as a chirality measure for spirals.

Our method is called SpArcFiRe, for SPiral ARC FInder and REporter. SpArcFiRe is both general and entirely automated, requiring only an image in which the galaxy in question is approximately centered. The method essentially "looks" at the image and uses computer vision techniques to produce a list of spiral arm segments. Although it is difficult to objectively and unambiguously define an "arm" (see Section 3.1), we produce a list of arm segments that would be plausible to a human observing the image. We provide the full specification of each segment found, including a list of pixels comprising the segment, its length, pitch angle, and all parameters of the logarithmic spiral (log-spiral) arc that best fit the cluster in a least-squares sense. Note that a segment may not always comprise a fully contiguous region of pixels, since our method will merge nearby segments when segment properties (close proximity and compatible pitch angles) allow it. The resulting segments are independent of each other, do not need to conform to any symmetry criterion, do not need to be attached to a bar or the bulge, and do not need to wind in the same direction.

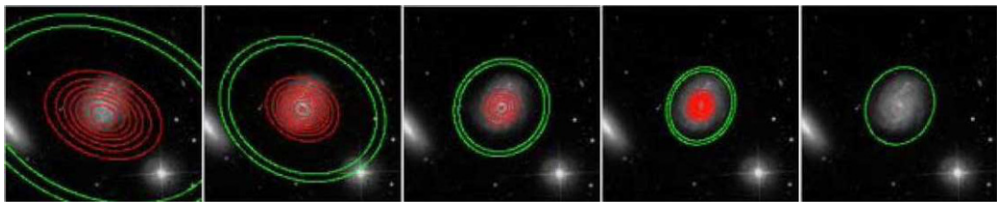

Figure 1. Finding and centering an approximately centered galaxy disk using successive 2D Gaussian ellipses. Assuming the galactic disk is circular when viewed face-on, the final ellipse (green, far right) allows us to estimate the angle at which the disk is viewed (modulo sign, which cannot be determined). This allows us to de-project the image to face-on for the ensuing steps of the algorithm.

(A color version of this figure is available in the online journal.)

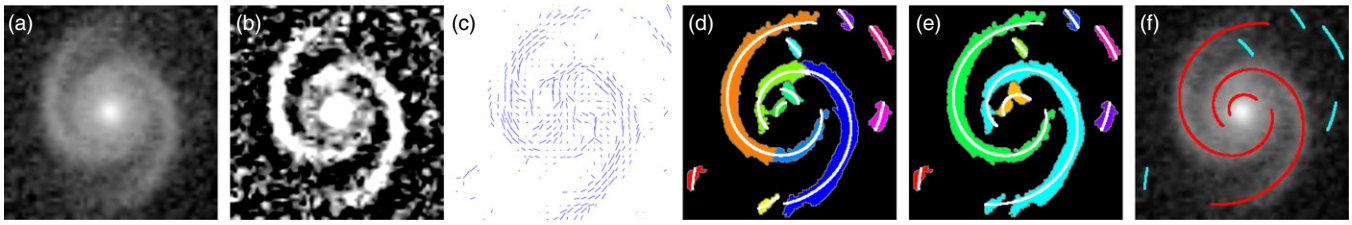

Figure 2. Steps in describing a spiral galaxy image. (a) The centered and de-projected image. (b) Contrast-enhanced image. (c) Orientation field (at reduced resolution for display purposes). (d) Initial arm segments found via Hierarchical Agglomerative Clustering of nearby pixels with similar orientations and consistent logarithmic spiral (log-spiral) shape, overlaid with the associated log-spiral arcs fitted to these clusters. (e) Final pixel clusters (and associated arcs) found by merging compatible arcs. (f) Final arcs superimposed on the image (a). Red arcs wind S-wise, cyan Z-wise.

(A color version of this figure is available in the online journal.)

So far as we are aware, this list of arm segments comprises the most detailed, fully automated, quantitative description of spiral arm structures currently available. Given the list of pixels for each segment, astronomers could easily perform whatever measurement of that segment they may wish, such as color, luminosity, brightness profile, and so on.

2. METHOD

2.1. Brief Description

The method is described in detail elsewhere (Davis & Hayes 2012, but see also Appendix A for algorithmic improvements since then). Here we provide only a brief description. Given an image (FITS, PNG, or JPG) with a galaxy approximately in the center, we use an iterative 2D Gaussian fit to find, exactly center, and estimate the size of the galactic disk (Figure 1). Using that fit, and making the simplifying assumption that the disk of the galaxy would be circular if viewed face-on, we rotate the image so that the long axis is vertical, and then linearly de-project it to reconstruct a face-on view of the galaxy (Figure 2(a)). After applying a contrast-enhancement filter based on an unsharp mask (Figure 2(b)), orientation-sensitive filters (Au 2006) are used to assign an orientation (strength and direction) to each pixel in the image. Essentially, an orientation is like a vector without an arrowhead. If, for example, there is an approximately horizontal "line" of bright pixels in the image with darkness on each side, then the pixels along the line would be assigned a strong horizontal orientation. The resulting orientation field is depicted in Figure 2(c). Pixels are then clustered into regions with locally similar orientations and consistent log-spiral shape. Figure 2(d) shows the resulting pixel clusters, with each being a different color. We emphasize that brightness plays no explicit role in this clustering, although it plays an implicit role through the creation of the orientation field. In particular, the border of a cluster of pixels is not directly based upon the edge of a bright patch, but upon the pixels outside the cluster having an orientation incompatible with those inside the cluster—although, indirectly, that orientation difference is ultimately based on brightness differences. Figure 2(d) also depicts a log-spiral arc associated with each cluster. The parameters for each arc are determined by a least-squares fit to the pixels in the cluster; the fit can be PNG intensity-weighted, if desired.

Sometimes, the requirement for consistent log-spiral structure will block the merging of two clusters that, in retrospect, "should" have been merged. In particular, as the clusters grow into their final shape, the arc fits may become more compatible than they were earlier. Thus, a second stage of merging is performed, based primarily upon compatible spiral arc parameters (Figure 2(e)). We note that, as depicted by the cyan arm segment, we allow merges between clusters that are not *exactly* adjacent, as long as they are close and are compatible in log-spiral structure. This allows us to join arms that have been partially obscured by dust lanes.[1] Figure 2(f) depicts the resulting arcs overlaid on the original de-projected image.

Although bar detection is not a primary focus of our algorithm, strong bars can sometimes be mistaken for arms, and so we attempt distinguish bar-containing clusters from arm-containing clusters. Although the bar detection works reasonably well in this capacity, it cannot find all bars, and so we do not recommend our code for reliable bar detection in other contexts. Bar detection takes place in several steps. After generating the orientation field and before the clustering, we attempt to detect if there is a (prominent) bar candidate. This uses two Hough transforms (a simple line detection method; see Duda & Hart 1972)—one using orientation and the other using brightness. If a bar candidate is detected, then during the clustering, a cluster's fit error is the minimum of the log-spiral fit error and bar fit error (with all bar parameters fixed). At the end of both clustering steps, if the bar is a better fit than a log-spiral arc for a particular cluster, the arc is replaced by the bar (if the bar is a better fit to more than one cluster, those clusters are merged).

[1] For example, the dark red arm in Figure 4(d) depicts the merging of segments 2 and 3 from Figure 4(b), which are clearly separated by a dust lane.

Table 1
Astronomically Interesting Outputs, per Arc

Column Name	Variants/Comments
Galaxy ID	Given for every arc; echo of user-provided name
Arc length rank	Integer, 1 = longest
Arc length	In pixels
Pitch angle	Positive for S-wise, negative for Z-wise
θ start	Start of visible arc
θ end	End of visible arc
Initial radius	Distance (in pixels) of arc from center at θ start (outer end for S-wise, inner end for Z-wise)
Num pixels	In the entire 2D cluster
Error per length	From least-squares fit
Error per pixel	From least-squares fit
Mean intensity	From PNG/JPEG—not physically meaningful
PNG partition color	Cluster's color in the PNG partition image
Average width	Defined as (num pixels)/(arc length)
Pitch angle error bar	Defined as arctan(width/(arc length))/2

Notes. Currently we normalize all images to 256 × 256 pixels. This means the half-width of the image is 128 pixels, with the visible disk being about 100 pixels in radius. Each pixel is thus about 1/100 radii across. We also output PNG images for each stage depicted in Figure 2, and a PNG image partitioning the clusters by color (like Figure 2(e) but without the white arcs).

Table 2
Astronomically Interesting Outputs, per Galaxy

Column Name	Variants/Comments
Galaxy ID	Echo of user-provided name
Disk axis ratio	For fitted ellipse; used to de-project
Chirality	Longest arc; majority; length weighted
No. of chirality votes	Two integers, one for each direction
Chirality, length-weighted vote	Total arc length for each chirality
Chirality agreement	Longest two arcs long enough, and agree?
Pitch angle	Of longest arc; per-arc average; length-weighted average; length-weighted avg. including only arcs of dominant chirality; PNG brightness-weighted.
List of pitch angles	Sorted longest-to-shortest
List of pitch angles agreeing with dominant chirality	Sorted longest-to-shortest
Bar score	Candidate and final (not too accurate)
No. of arm segments	As a function of minimum length, for several length thresholds (useful as one measure of arm count, without claiming it's "the" arm count)

Finally, we note that the code has many user-changeable parameters, although we do not expect that the parameters will need to be changed on a per-galaxy basis, but instead on an image-set basis. For example, we use one set of parameters for the entire Sloan set of galaxies. User-changeable parameters include the amount and radius of the unsharp mask, the threshold fit error ratio for allowing cluster merges, the threshold distance under which non-contiguous cluster merges are allowed, the size threshold below which cluster merges are not checked for log-spiral arc compatibility, etc. Such parameters and their effects, as well as how our code responds to decreasing image resolution, will be discussed in detail in a forthcoming paper.

2.2. Algorithmic Outputs

When run on a set of galaxies our algorithm produces two tables in CSV format: the first (Table 1) provides a detailed list of outputs on a per-arc basis, so that each galaxy is listed several times (once for each arc it contains). The second provides a per-galaxy summary (Table 2).

Some of the per-arc values in Table 1 warrant discussion. Each spiral arm segment is defined first and foremost by the list of pixels comprising it. This list is not provided in the CSV output, although we provide a sequence of PNG images for each step of the algorithm, one of which includes a partition of the PNG image into clusters giving arm-segment membership by color, with the colors specified in the CSV file. We then perform a least-squares fit of a log-spiral arc (see Davis & Hayes 2012 for details) to this cluster of pixels, with the start and end of the visible arc simply determined according to the minimum and maximum polar coordinate θ values for that cluster. The total least-squares error is calculated and then divided by either the arc length or number of pixels—we are unsure which is a more meaningful measure of the fit error for astronomical purposes. For each galaxy, the resulting arcs are listed in decreasing order of length, identified by an integer starting at 1 (the "arc length rank"); this rank also specifies the order in the list of pitch angles in Table 2. We also provide the mean intensity (per pixel) of the cluster, although this measure is taken from the PNG/JPEG images, not from the original FITS image. (If the latter is desired, the user must re-project the cluster back onto the FITS image and perform the measurement themselves.) We define the average width of the arm segment by comparing it to a "rectangle" (in polar coordinates) with the same length and number of pixels as the cluster. The width of this imaginary rectangle is (pixel count)/(arc length), and so we define that as the average width of the arm segment. This average width also suggests a definition of an error bar for the pitch angle, defined as half the angle subtended between the diagonal across the rectangle and the longitudinal side of the rectangle.

Some of the per-galaxy values in Table 2 have more than one variant. For example, the chirality is always determined from the sign of the pitch angle of the arm segments, but the determination can be different depending upon how we combine the arc votes. As we will see in the next section, the one-vote-per-arc voting scheme does not do as well as other measures; an arc-length-weighted vote does better (which is not surprising as longer arms indicate a stronger signal), although even the chirality of the single longest arc does quite well. The most complicated value is the average pitch angle of the galaxy, which can be determined in myriad ways: the longest arc alone; the average of all of the arcs, regardless of length or sign; a length-weighted average; a (PNG/JPEG) brightness-weighted average; and finally, a length-weighted average based only upon those arcs whose chirality agrees with the length-weighted dominant chirality. We believe the last option is the best candidate for the pitch angle of a galaxy, if one number is to be assigned to the pitch angle of the galaxy. Such a pitch angle could be compared with models that assume only one pitch angle per galaxy; such models have suggested that pitch angle may correlate with the mass of the bulge and/or central black hole (Seigar et al. 2008; Ringermacher & Mead 2009). We also

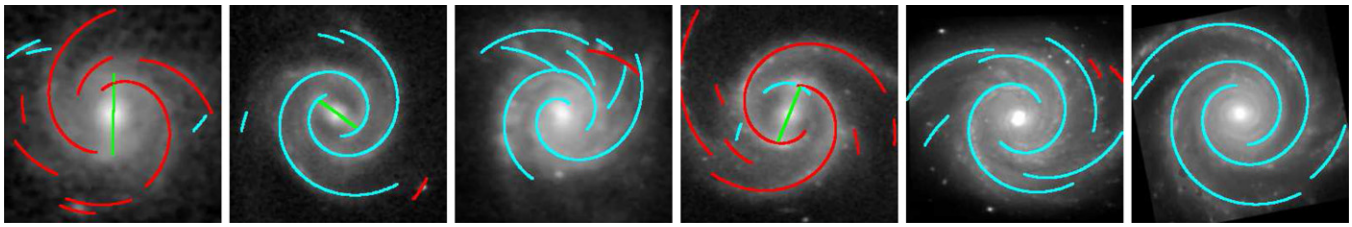

Figure 3. Typical examples of how the algorithm performs on "nice" galaxies. Green represents a bar, red arms wind S-wise, cyan arms wind Z-wise. Per-galaxy outputs are listed in Table 3. The galaxy names and parameters of all the arcs are listed in Table 4.

(A color version of this figure is available in the online journal.)

Table 3
A Subset of the Galaxy-level Output for the Six Galaxies in Figure 3

ID	MA	AR	*A*	*L*	2	*B*	*L*	*R*	PAL	PAM	PAA	PSD	AAL	MAL	A50	R50	#⩾0	40	100	200	300	500
1	92	0.85	S	S	T	T	47	F	18.8	18.8	17.6	5.0	81.5	60.0	110	3	11	8	3	1	0	0
2	197	0.92	Z	Z	T	T	27	T	−11.9	−11.9	−13.0	2.3	160.1	30.0	401	1	5	2	2	2	2	0
3	143	0.77	Z	Z	T	F	46	F	−19.3	−19.3	−20.2	12.5	121.2	71.5	270	2	8	7	3	2	0	0
4	229	0.85	S	S	T	T	34	T	15.5	17.8	17.6	4.6	108.4	40.0	264.0	2	9	5	3	2	1	0
5	281	0.94	Z	Z	T	F	16	T	−13.9	−13.9	−16.1	9.3	129.0	132.0	247	2	8	6	5	2	0	0
6	269	0.99	Z	Z	T	F	103	F	−10.4	−10.4	−10.5	2.4	285.5	272.5	470	2	4	3	2	2	2	1

Notes. Their names are given in Table 4. Legend: we split the columns into sections as follows: *ID*. normally user-supplied, but for the purposes of this table, the left-to-right number of the galaxy from Figure 3; *Size and tilt of the disk*. MA: disk major axis length in native pixels from ellipse fit (see Figure 1); AR: disk axis ratio (corresponds to disk tilt from line-of-sight—note images in Figure 3 are de-projected); *Chirality*. *A*: arc-length weighted; *L*: chirality of the longest arc; 2: do the two longest arcs agree in chirality? *Bar/central region*. *B*: is there a bar candidate? *L*: bar half-length in standardized pixels (modeled even if it was not used; see previous column); *R*: was the central cluster removed from spiral arc consideration? This happens if it is best modeled by a bar, or if it contains the central pixel even in the absence of a bar candidate. See, for example, Galaxy 1 above: the central area is *not* removed from spiral arc consideration even though there is a bar, either because the central region was not *best* modeled by a bar, or because the central cluster did not contain the central pixel. *Pitch angle*. PAL: pitch angle of the single longest arc. The following pitch angles are all length weighted, considering only arcs agreeing with the dominant chirality—PAM: median pitch angle; PAA: average pitch angle; PSD: standard deviation of pitch angle; *Arcs*. AAL: average arc length; MAL: median arc length; A50: arc length at 50% of total length (see the text); R50: rank at 50% of total length (see the text). Finally, the last set of columns lists the number of arcs as a function of minimum length in standardized pixels, which are pixels when the image has been rescaled to 256 × 256.

recommend that our error bar be included with the pitch angle; this error bar is based on a length-weighted standard deviation of the pitch angles across arcs (although whether it should be across *all* arcs or only those of dominant chirality is uncertain, so we provide both). Surprisingly, initial tests on the (PNG/JPEG) brightness-weighted average did not seem promising, possibly because (1) bright knots can cause a segment to be weighted too strongly, and (2) dim but well-defined arms should not be ignored. Determining a better method of brightness-based weighting may be an area of further work.

Although the notion of arm count is ambiguous (because the concept of a "spiral arm" is not well defined), we include several basic measures that could provide a plausible arm count. These measures count the number of arcs exceeding a variety of length thresholds. These counts are included across all arcs, as well as only across the arcs that agree with the arc-length-weighted dominant winding direction. A comparison of these arc counts against human arm counts is presented in Section 4.2.3.

Figure 3 shows some typical "nice" examples. In general, we find that if the image is clear and has a sufficient signal-to-noise ratio (about as good as needed by a human, but no more), then the algorithm does a very good job of marking out the arms and determining pitch angle. However, we believe that an important strength of our algorithm is that it *also* does well when the galaxy is less clean, less symmetrical, and more fragmented. Even when the image has such low resolution that little structure is visible, we find good agreement with human determinations of structure.

Table 3 lists a selection of the per-galaxy outputs for the galaxies depicted in Figure 3. (It is only a selection because the per-galaxy output contains upward of 120 columns, many of which are computable from the per-arc table.) The two columns A50 and R50 deserve extra discussion. They are, respectively, the arc-length and arcrank of the arc that lies at 50% of the total length of all arcs when all the arcs are laid end to end, longest to shortest. A50 is an interesting measure because it provides a sense of how much of the total arc length resides in long versus short arcs. For example, if most of the arcs are long, then when laid out end to end, longest to shortest, the arc that resides halfway along the entire length will be a long arc. In contrast, if there are a multitude of short arcs among a few long ones, then the length of the arc at the 50% point may be rather short, even if the mean length is quite long. Although it may seem that this measure is similar to the median arc length, it differs in that it is essentially a length-weighted median, rather than a median among the *list* of arcs. Observe, for example, that in all cases the A50 measure is significantly longer than both the mean and median arc length; this is a property precisely of the nice galaxies that we chose to display—namely, that nice galaxies tend to have very long arms. R50 is similar to A50; it is the rank (sorted by length) of the arc containing the 50% length position. Again, it is different from the median rank (which would just be the arc rank listed in the center row (for each galaxy) in Table 4); instead it is a length-weighted median rank.[2]

[2] This measure is included because a similar measure was very useful when the second author was doing research in genome-sequence assembly, where long assembled sequences were of much higher value than short ones. We found this measure to be highly valuable in measuring the quality of a proposed genome assembly because it gave an indication of how much the total assembled sequence was in long strings rather than short ones.

Table 4
Arcs of the "Nice" Galaxies from Figure 3

N	Galaxy Name/SloanID	ArcLengthRank	Length	NumPix	PitchAngle	θ_S	θ_E	R_S	R_E	Err/L	Err/Pixel	I_μ
1	1237660635996291172	1	230	3678	18.8	1.78	4.50	48	123	154.5	9.68	0.332
1	1237660635996291172	2	188	3103	18.7	1.56	4.72	31	92	470.9	28.61	0.403
1	1237660635996291172	3	110	1393	23.7	2.59	3.69	71	115	80.7	6.42	0.293
1	1237660635996291172	4	77	641	7.9	2.78	3.50	102	112	27.9	3.38	0.206
1	1237660635996291172	5	63	946	21.7	2.87	3.42	96	119	89.3	6.03	0.235
1	1237660635996291172	6	60	824	18.1	2.63	3.65	48	67	115.0	8.51	0.426
1	1237660635996291172	7	42	398	−25.1	3.00	3.29	123	141	33.1	3.51	0.239
1	1237660635996291172	8	42	637	7.3	2.97	3.31	118	123	108.8	7.19	0.253
1	1237660635996291172	9	29	159	8.2	3.01	3.28	104	109	4.5	0.84	0.221
1	1237660635996291172	10	28	362	−38.1	3.05	3.24	107	124	108.6	8.48	0.264
1	1237660635996291172	11	21	167	−34.6	3.05	3.23	95	108	17.9	2.33	0.231
2	1237661958829441105	1	401	6232	−11.9	−4.84	2.00	25	108	229.3	14.76	0.366
2	1237661958829441105	2	316	5213	−13.1	0.31	5.98	26	97	236.4	14.37	0.372
2	1237661958829441105	3	30	246	−21.9	2.99	3.30	84	96	17.6	2.18	0.212
2	1237661958829441105	4	30	447	9.8	3.03	3.25	128	134	66.0	4.46	0.175
2	1237661958829441105	5	21	154	−20.2	3.05	3.23	107	115	7.5	1.07	0.163
3	1237662238014308394	1	293	4630	−19.3	0.19	6.10	13	111	717.3	45.41	0.495
3	1237662238014308394	2	270	4868	−10.8	0.72	5.56	33	84	775.4	43.15	0.487
3	1237662238014308394	3	134	2064	−14.3	2.51	3.77	87	120	317.7	20.68	0.321
3	1237662238014308394	4	75	960	−49.5	2.81	3.47	49	106	390.4	30.56	0.422
3	1237662238014308394	5	68	638	18.2	2.78	3.50	80	101	38.6	4.12	0.338
3	1237662238014308394	6	55	600	−46.8	2.88	3.41	53	93	129.6	11.94	0.412
3	1237662238014308394	7	51	515	−26.1	2.77	3.52	50	73	73.6	7.32	0.467
3	1237662238014308394	8	21	274	−3.2	3.03	3.25	95	96	78.6	6.14	0.310
4	1237667783367917730	1	329	5872	15.5	0.52	5.76	26	114	360.0	20.23	0.430
4	1237667783367917730	2	264	4894	17.8	1.25	5.03	34	114	364.0	19.65	0.416
4	1237667783367917730	3	182	2115	18.8	2.48	3.80	103	162	74.3	6.39	0.253
4	1237667783367917730	4	59	710	−5.4	2.32	3.97	33	38	129.3	10.86	0.574
4	1237667783367917730	5	40	471	11.7	2.88	3.40	72	81	58.6	5.07	0.356
4	1237667783367917730	6	29	211	10.7	3.03	3.26	123	128	7.2	1.01	0.178
4	1237667783367917730	7	27	266	32.5	3.01	3.28	78	93	53.2	5.52	0.289
4	1237667783367917730	8	23	200	36.7	3.03	3.25	78	92	33.5	3.86	0.281
4	1237667783367917730	9	18	167	−35.2	3.02	3.26	58	69	64.6	7.24	0.397
5	NGC 4321	1	287	5276	−13.9	0.77	5.52	30	99	580.6	31.69	0.462
5	NGC 4321	2	247	4155	−10.8	0.93	5.36	35	81	304.9	18.18	0.520
5	NGC 4321	3	139	1826	−25.0	2.46	3.83	66	125	147.8	11.30	0.362
5	NGC 4321	4	132	2000	−23.1	2.59	3.70	85	137	180.8	11.94	0.314
5	NGC 4321	5	132	1503	−4.4	2.44	3.84	89	99	134.2	11.79	0.319
5	NGC 4321	6	45	497	−46.6	2.98	3.30	83	116	117.2	10.64	0.276
5	NGC 4321	7	24	255	0.8	3.03	3.25	107	108	41.1	3.90	0.276
5	NGC 4321	8	23	159	15.5	3.05	3.23	123	130	11.2	1.65	0.241
6	NGC 4939	1	563	8433	−10.4	−5.03	2.55	33	135	261.8	17.48	0.395
6	NGC 4939	2	470	7167	−9.8	−5.29	1.39	36	116	193.7	12.72	0.402
6	NGC 4939	3	75	1180	−9.3	2.83	3.46	111	124	148.3	9.44	0.296
6	NGC 4939	4	33	157	−23.9	3.01	3.27	111	125	2.6	0.56	0.218

Notes. (θ_S, R_S) and (θ_E, R_E) are the start and end points of the arc in polar coordinates (angles in radians), in the standardized image; err/L and err/pixel are the least-squares fit error of the arc to the pixels in the cluster per unit length and per pixel, respectively; and I_μ is the mean intensity of the pixels in the cluster, as defined in the PNG image, which is not necessarily related to a physical intensity. All lengths are in pixels in the standardized image (256 × 256).

Table 4 lists all the arc-level output for the galaxies in Figure 3, and should be relatively self-explanatory.

3. SOME PRELIMINARY OBSERVATIONS

3.1. Defining an "Arm" is Hard

Astronomers frequently refer to spiral arms in a galaxy, but to our knowledge a formal definition of what, exactly, constitutes a spiral arm does not exist. Part of the reason is probably that defining what constitutes a visual arm is not easy. (On a related note, we shall see in Section 4.2.3 (see Figure 13) that humans also find it difficult to agree on how many arms are in a spiral galaxy.)

Because our list-of-arcs representation produces well-defined arcs, but the notion of a spiral arm is not well defined, we will refer to the former as an arc, and refer to the region of the arm described by the arc as a "spiral arm segment." We will reserve the term "spiral arm" for the (sometimes ambiguous) reference to the physical phenomenon in spiral galaxies. Figure 4 illustrates the need for this distinction: arm segments can fork or split, have dust lanes, and have kinks in them where the pitch angle changes even if the arm segments are visually contiguous. As a result, an arm segment is defined as an (almost)

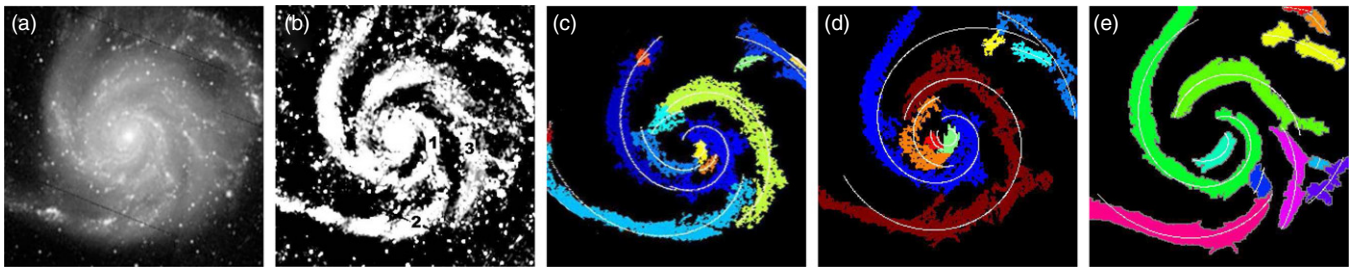

Figure 4. Three different interpretations of the spiral structure in M101. The colored images are from different versions of the code. All comprise reasonable interpretations of the structure. (a) Original image. (b) Contrast enhanced. We have labeled three arm segments; the joint between segments 1 and 2 may be what a human would call a fork, although our code never refers to forks. (c) An old version of the code, where the three segments happen to be separated (blue, cyan, and olive pixel clusters). (d) An intermediate version of the code, where segment 2 has been interpreted as a continuation of segment 3, jumping over the gap between them. The single log-spiral arc spanning the two fits reasonably well, which suggests that perhaps segments 2 and 3 are physically one arm with an obscuring dust lane, while the apparent fork of segment 2 from segment 1 is an optical illusion. (Note: the arc at the top of this figure was not fitted correctly because arcs spanning more than 2π radians were not always properly handled at that time; they are now handled correctly.) (e) The most recent version of the code, in which the log-spiral arcs are more stringently fit (in the least-squares sense) to each cluster of pixels.

(A color version of this figure is available in the online journal.)

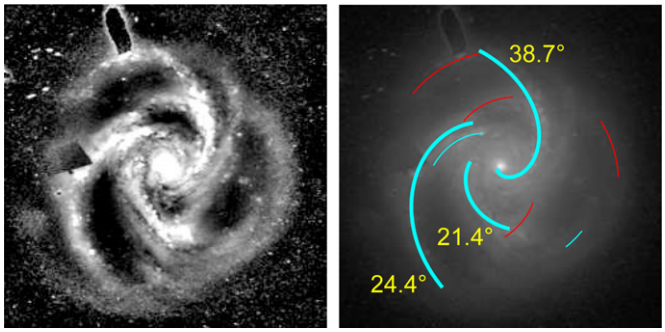

Figure 5. NGC 5054 has arms of widely varying pitch angle. The two longest arcs differ in pitch angle by 14.°3; the pitch angle of the upper arc remains about the same (37.°8) even if we exclude the near-center region (a discontinuous region that was joined during the secondary merging step).

(A color version of this figure is available in the online journal.)

contiguous region of light with a locally consistent, smoothly changing orientation that traces out an arc that can be closely approximated by a log-spiral.

We note that the distinction between spiral arm and spiral arm segment does not indicate a loss of information; it is an artifact of describing an ambiguous concept with precise quantitative information.

3.2. Pitch Angles Can Vary Significantly between Long Arm Segments in the Same Galaxy

Observing the arcs superimposed on the images in Figures 2–4, we make two observations. First, the log-spiral arc is a very good mathematical description of the curve of the spiral arm segments. Second, even though the fit is usually very good, the pitch angle between different arms in the same galaxy can be quite different. For example, Table 4 lists the parameters of the arcs from the galaxies in Figure 3; the pitch angle differences between the two longest arcs are more than 8°. Figure 5 provides a stark example in NGC 5054, which has long arms that differ in pitch angle by more than 15°. Our code demands that all arms are fit quite well, so the segments have a width much smaller than their length and the arcs mostly remain inside their cluster for their entire length (e.g., Figure 2(e)). This indicates that the error in these pitch angles is much less than their difference. Such a disagreement is not atypical, as Table 5 shows. Recent work has also suggested that pitch angle can also vary with radius, and that this variance correlates with other

Table 5
Median Pitch Angle Difference Between the Two Longest Arms in a Galaxy (Bottom Row) When Both are at Least as Long (in Pixels) as the Top Row

Minimum length	0	50	100	150	200	250	300
Median difference	14.5	14.3	10.7	7.5	5.6	3.5	2.6

Note. Note that the images are scaled to 256 × 256 pixels.

structural parameters of the galaxy (Savchenko & Reshetnikov 2013). Figure 6 demonstrates that average pitch angle increases with radius over our set of 29,250 spiral galaxies (described in footnote 4). All images are scaled to 256 × 256, which means that the disk radius is about 100 pixels. Note that it may be hard to detect this progression in any individual galaxy because (1) we model each individual spiral arc as having a constant pitch angle, and (2) pitch angles can differ significantly between arcs in one galaxy. Despite this being very raw data without any interpretation or filtering, there is quite a strong signal indicating a correlation between average pitch angle and radius.

4. COMPARISON WITH PREVIOUS WORK

So far as we are aware, the current work is the first large-scale, automated, quantitative survey of detailed general spiral structure in galaxies. There exist several small-scale quantitative surveys (e.g., Ma 2001; Davis et al. 2012, no relation), and two large-scale, qualitative, human-based surveys: one citizen science project—the Galaxy Zoo (GZ; Lintott et al. 2008, 2011; Willett et al. 2013)—and one professional survey (Nair & Abraham 2010). We compare our results to all except the last, which includes many interesting measures but none we could easily compare to our own.

4.1. Quantitative Comparison with Existing Small Surveys

We compare our results to two small, quantitative surveys. In the first, Ma (2001) manually measured the pitch angles of either one or two arms in each of a small sample of galaxy images; we have downloaded and run our algorithm on the images used in that paper. The second group, based in Arkansas, uses a Fourier analysis to extract the dominant pitch angle (Seigar et al. 2006; Davis et al. 2012); we downloaded and ran our algorithm (including star masking) from the Carnegie–Irvine Galaxy Survey (Ho et al. 2011). In Figure 7, we compare our results to both of these studies. As can be seen, the

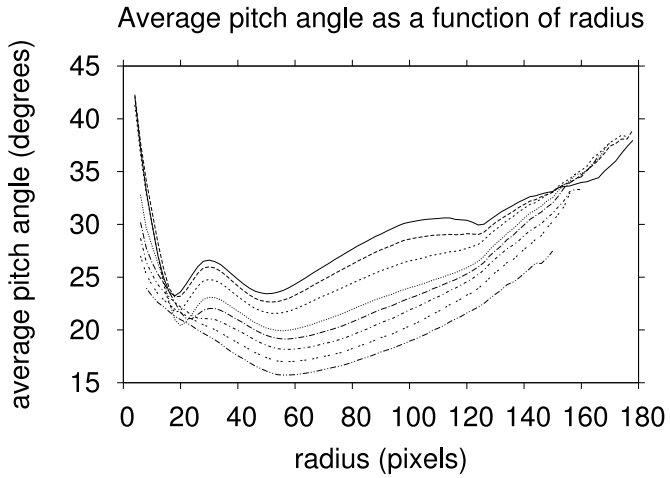

Figure 6. Plot of average pitch angle as a function of radius, across all arcs in all galaxies in our set of 29,250 spiral galaxies (see footnote[4]). Each curve represents the outcome of restricting arcs to having a minimum length; labeling the curves from top to bottom (at $r = 100$) they represent arcs of length at least 25, 50, 75, 100, 125, 150, 175, and 200 pixels, respectively. For each value of radius r, the average pitch angle is computed across all arcs (independent of galaxy, and having minimum length as specified) that touch that value of r. (In other words, an arc is included in the average at r if it is long enough and satisfies $R_S \leqslant r \leqslant R_E$—see Table 4.) A value is plotted at r only if there are at least 100 arcs touching that radius. This is very raw data and some of the observed effects may be spurious; for example, the high pitch angles below $r = 20$ could be contaminated by undetected bars, and the dip near $r = 128$ is likely due to the edge effects because our images are 256×256, meaning that the nearest edge is 128 pixels from the center. Note that the corners of all our images are $128\sqrt{2} \approx 181$ pixels from the center, meaning that the largest possible value of r is 181. The trend of decreasing pitch angle with increasing arc length may simply be due to the fact that, in an unperturbed galaxy, longer arcs need to wind more tightly in order to squeeze into a disk of fixed radius; it is unclear if there is a deeper physical effect occurring. The hump near $r = 30$ is a mystery, although one may hazard a guess that the dips on either side may be due to short ring-like segments sometimes seen at the end of a bar.

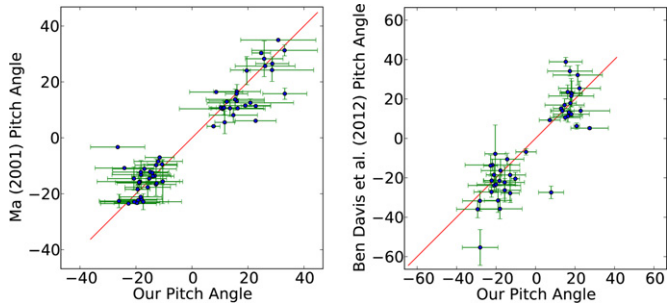

Figure 7. Scatter-plot comparison of our pitch angles to other sources. Left: with Ma (2001). Right: with Seigar et al. (2006) and Davis et al. (2012), both located in Arkansas. In both cases, the vertical axis is the pitch angle measured by the other authors, and the horizontal axis is our measured pitch angle. Our error intervals are derived from the length-weighted standard deviation of pitch angle across arcs. Intervals for the Arkansas group are from their corresponding publications; for Ma (2001), we derived the intervals from the difference between the two measured arcs (when two measurements were available).

(A color version of this figure is available in the online journal.)

results always agree in chirality (with one exception, where many foreground stars were present, forcing there to be many star-cleaning artifacts),[3] although there is some scatter in the measured pitch angles. This scatter may seem to be a concern, however recall from Table 5 and Figure 5 that the pitch angle can

[3] Note that the Carnegie-Irvine Galaxy Survey (CGS) images used PNGs with CGS star cleaning, whereas the other image sets used FITS with our own star removal. We do not expect this to significantly alter our results.

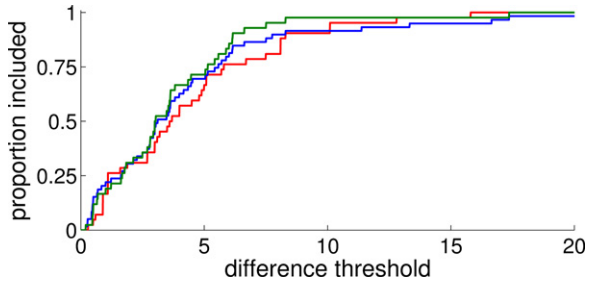

Figure 8. Cumulative distribution of pitch angle discrepancies between the two arcs measured for each galaxy from (Ma 2001; red), and between our measurements and the measurements in (Ma 2001, blue for the full comparison set and green for the subset where two arc measurements are available from Ma 2001). Since the curves are all very similar, it is plausible that much of the scatter between our results and the results of other authors arises from genuine within-galaxy arm variation rather than between-method measurement variation.

(A color version of this figure is available in the online journal.)

vary quite significantly between arms in a single spiral galaxy. Because Ma (2001) measured only one or two arms in each galaxy and the Arkansas group measured only one dominant pitch angle, it is possible that the scatter could be explained by inter-arm differences in each galaxy rather than differences as a function of method. To test this hypothesis, Figure 8 plots the cumulative distribution of pitch angle discrepancies between the two arcs in one galaxy (when available) from Ma (2001) versus discrepancies between Ma's method and our method. All three curves are very similar, so it is likely that much of the scatter in Figure 7 arises from within-galaxy arm variation rather than between-method measurement variation.

4.2. Statistical Comparison with Large Human Surveys

4.2.1. Chirality Comparisons

Galaxy Zoo (Lintott et al. 2008, 2011) is a citizen science project in which approximately 250,000 human volunteers classify images of galaxies over the Web after some rudimentary training. The median number of people who viewed each image was about 40, which means that some measure of certainty can be obtained from multiple viewings of each image. Galaxy Zoo 1 (GZ1; Lintott et al. 2011) presented people with an image of a real galaxy along with six cartoon galaxies, and asked them to choose which cartoon most resembled the real galaxy. Across spiral galaxies with observable structures, the only comparison we can make with GZ1 is chirality (S-wise versus Z-wise winding direction). In difficult cases some humans may choose spiral while others do not, so we compare our results on chirality against what we call the discernibility of a galaxy: the maximum fraction of agreeing humans that saw spiral structure, which we define for a particular galaxy as (max(S-wise votes, Z-wise votes)/total number of votes). For example, a discernibility of 60% indicates that 60% of humans voted for one chirality, whereas the other 40% voted either for the other chirality, or for some non-spiral classification. A discernibility of 100% means that all human observers clearly saw spiral arms and they all agreed on the chirality. Note that discernibility can be less than 50%, because there were six choices in GZ1, but can still be meaningful if S-wise or Z-wise was the most popular vote among the six choices.

Table 6 compares our chirality measurement against those of GZ1 humans as a function of both human discernibility and several measures of chirality derived from our output. We

Table 6
Winding-direction Agreement with Human Classifications from Galaxy Zoo 1

Min discernibility rate	0	40	60	80	90	95	100	0	40	60	80	90	95	100
Our longest 2 arcs agree?	*F*	*F*	*F*	*F*	*F*	*F*	*F*	*T*	*T*	*T*	*T*	*T*	*T*	*T*
Inclusion rate	100.0	98.9	95.7	79.4	52.5	32.1	12.5	74.4	73.9	72.1	61.9	43.2	27.6	10.9
Mean discernibility	87.1	87.7	88.9	92.2	95.9	98.0	100.0	88.6	89.0	90.0	92.8	96.1	98.0	100.0
Majority vote	80.6	80.8	81.1	82.0	83.8	85.5	86.5	83.9	84.0	84.2	84.8	86.0	87.2	87.9
Longest arc alone	89.3	89.4	89.8	91.2	93.2	94.8	95.8	97.6	97.7	97.9	98.4	98.9	99.3	99.6
Length-weighted vote	93.3	93.4	93.7	94.7	96.4	97.5	98.2	97.4	97.5	97.7	98.1	98.8	99.2	99.5

Notes. Row 1: first independent variable: the minimum proportion of human votes (which we call the "human discernibility") that the majority-vote winding direction must receive among the six GZ1 categories. Row 2: second independent variable: do we demand that our two longest arcs agree in chirality? (Both must be at least 64 pixels long.) Row 3: the proportion of the 29,250 galaxies included under the above criteria. Row 4: the *mean* discernibility rate among the galaxies included under these criteria. Rows 5–7: agreement rates between Galaxy Zoo 1 and three methods of determining winding direction from our output.

include 29,250 galaxies with clearly observable structure as defined by GZ votes, chosen by the director of the GZ project (S. Bamford, 2011 private communication).[4] We see that the best measure is the "length-weighted vote," which outperforms all other measures (although "longest arc alone" is competitive); it agrees with humans at least 93.3% of the time, increasing to 98.2% of the time as human discernibility increases. If we additionally restrict ourselves to galaxies in which our longest two arcs agree, then the "length-weighted vote" agrees with humans between 97.4% and 99.5% of the time, depending upon human discernibility. (We note that "longest arc alone" slightly outperforms "length-weighted vote" here, but not by a significant amount.)

Another interesting question to ask is: How good is SpArcFiRe at determining chirality, compared to an average individual human? We can give an approximate answer to this, as follows. Given a set of galaxies S, let the mean discernibility across those galaxies be $\bar{D}$ and let our length-weighted vote agreement with the majority (as listed in the bottom row of Table 6) be $\bar{L}$. For the humans, this means that for the average galaxy in S, $\bar{D}$ percent of the voting humans agree with the majority human vote. Conversely, for the average galaxy in S, SpArcFiRe agrees with the human majority $\bar{L}$ percent of the time. Thus, the bottom row of Table 6 is directly comparable to the mean discernibility row. Assuming that the human majority vote is the "correct" answer, we see from Table 5 that, except for the highest discernibilities (95% and 100%), for an average galaxy our length-weighted vote (on average) agrees with the human majority more frequently than individual human voters do. Even in the highest discernibilities, SpArcFiRe agrees about 98% of the time with the human majority. Finally, when ordering by discernibility, more than half (the lower 56.56%) of the galaxies have an average agreement higher than their average discernibility, so in some sense we can say that, on average, SpArFiRe is more reliable than individual humans, at least in the simple case of determining chirality. (See Appendix B for further discussion and a derivation showing that the mean discernibility of galaxies is equivalent to the mean consistency among humans.)

In the most stringent cases where both humans and SpArcFiRe have high confidence (see the far right column of Table 6), SpArcFiRe agreement is so high with humans that there are only 15 galaxies in which our length-weighted vote disagrees with the humans. To see what went wrong, it is instructive to look more closely at some of these instances. First, there are four cases where the image standardization step (see Figure 1) zooms in too far, cutting off dim outer arms and resulting in the chirality being computed from arcs that are essentially noise. Second, we find eight cases in which the two longest arcs are spurious and the total length of spurious arcs is greater than that of the correct chirality arcs. Third, we find three cases in which the two longest arcs are of the correct chirality, but the total length of smaller, incorrect chirality arcs is longer than the total length of correct ones, resulting in the arc-length-weighted vote going in the wrong direction. (Some of the cases suffer to some extent from more than one of the stated problems.) Figure 9 depicts one case of each type of error. One may be tempted to try to adjust the algorithm to fix these problems. However, we have already spent a significant amount of time playing that game (years); we must remember that these cases comprise only a tiny percentage of the total test set, and most tweaks aimed at reducing errors in this set of 15 galaxies will likely introduce more problems elsewhere than they solve here.[5]

Longo (2011) also provided a chirality measurement for a survey of galaxies, in which each galaxy was viewed only once by one of five student volunteers. The students were told to choose a chirality only if it was clear. Table 7 lists our agreement with Longo across the intersection of our two samples. Comparing the two tables, we see that our agreement with Longo's individuals roughly matches the GZ1 agreement columns with minimum discernibilities of about 80%. We are not sure how to interpret this observation; perhaps it suggests that, on average, an individual human's judgment thinks something is clear, when a crowd of humans would be in about 80% agreement.

4.2.2. Pitch Angle Comparisons

We now compare pitch angle measurements with those of Galaxy Zoo 2 (GZ2; Willett et al. 2013). Here, where human classifiers indicated that there was "any sign of a spiral arm pattern," they were asked whether the spiral arms were tight, medium, or loose; the corresponding icons are depicted in Figure 10. Figure 11 shows the relationship between our measured pitch angle (arc-length-weighted vote) and the proportion of galaxies receiving a majority human vote for

[4] This selection is a carry-over from Davis & Hayes (2012), despite GZ2 data having become publicly available in the interim. Not all 29,250 galaxies could be found in the newest Sloan Digital Sky Survey (SDSS) release. We keep the old selection for consistency, and do not expect this selection to significantly alter our results.

[5] Trust us on this. The first author spent several years of his PhD fiddling with the algorithm to get to this point.

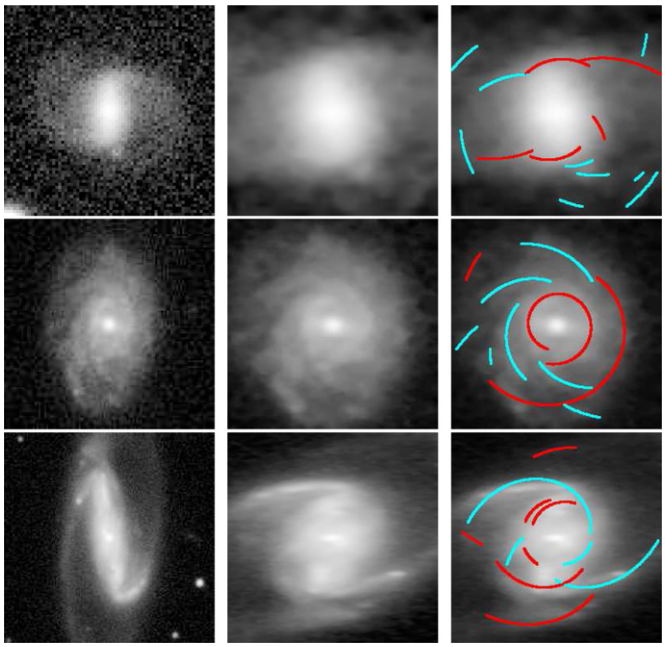

Figure 9. Three of the 15 pathological cases in which GZ1 humans are in 100% agreement on chirality, but our length-weighted vote gets it wrong even though our two longest arcs agree in chirality (see bottom right corner of Table 6). Top row: zooming in too far past very low surface brightness arms (four cases). Middle row: the two longest arcs (red) agree but are spurious (at least with respect to chirality) and outweigh arcs with correct chirality (eight cases). Bottom row: the two longest arcs (cyan) agree and have correct chirality, but are outweighed by numerous shorter arcs, not all of which are necessarily spurious—see for example the long red arcs at the bottom of the lowest image, which are arguably real arcs even though they disagree with the correct chirality (three cases).

(A color version of this figure is available in the online journal.)

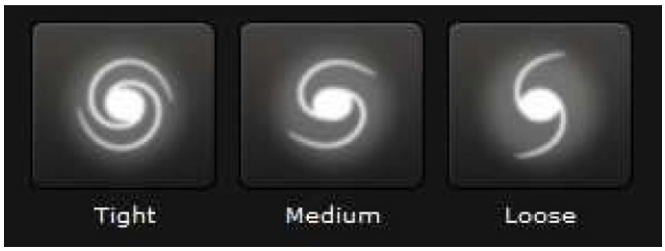

Figure 10. Reference icons presented to human classifiers when asked about spiral arm tightness during the Galaxy Zoo 2 project.

Table 7
Agreement between our Chirality and That of Longo (2011), for the 13,331 Galaxies Intersecting the Set of 29,250 Classified Confidently as Spiral by Galaxy Zoo Humans

	All Galaxies	Longest 2 arcs Agree
Majority vote	83.2	86.1
Longest arc alone	91.6	98.3
Length-weighted vote	95.1	98.1

Note. The last column uses the subset of galaxies where our two longest arcs agree and are at least 64 pixels long.

Tight, Medium, or Loose. In this and later comparisons, we calculate our pitch angles as an arc-length-weighted average of pitch angles of individual arcs, using the arcs that agree with the chirality indicated by the length-weighted sum of all arcs. As can be seen from the dashed lines in Figure 11, galaxies where we measure a low pitch angle usually have a majority of

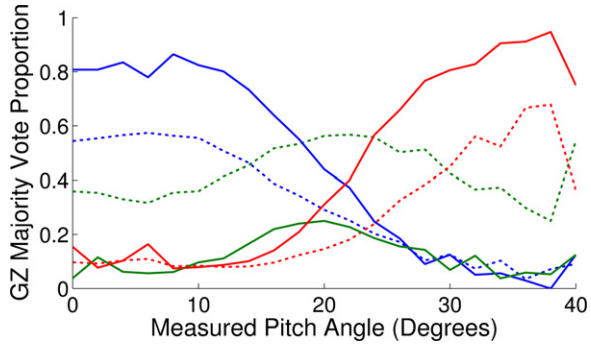

Figure 11. Proportion of galaxies receiving a majority vote for Tight (blue lines), Medium (green lines), or Loose (red lines) as a function of our measured pitch angle. Pitch angles are binned with width 2° between 0 and 40, with just one bin beyond 40 (due to low sample size). The dashed lines represent all images tested from Galaxy Zoo 2; the solid lines represent the images within the top quartile of human agreement.

(A color version of this figure is available in the online journal.)

human votes for Tight, whereas most of the remaining votes in this range were for Medium tightness. As our measured pitch angle increases, we see progressively fewer galaxies classified as Tight, and more galaxies classified as Loose. Designations as Medium are pervasive throughout, and Loose classifications are less frequent than Tight. This reflects the classification distribution of the image set as a whole.

In the top quartile of human agreement (lowest Shannon entropy quartile), the association between our tightness measure and human classifications is even more pronounced, as shown in the solid lines of Figure 11. Also, as the Medium-majority votes were far less common with increased human agreement, it seems likely that this choice was frequently used to indicate uncertainty, perhaps due to low galaxy resolution or galaxies with arms of different tightness. Consequently, it is reasonable that Medium-majority galaxies spread across a wide range of our measured pitch angles. Even if Medium-majority galaxies are disfavored by the entropy measure from having two neighbors (despite most galaxies likely appearing closer to Tight or Loose), such willingness to put many galaxies in any of the three categories would further suggest that much of the spread in classification stems from human uncertainty. In all, then, we see a clear association between GZ tightness classifications and our measurements, with this association strengthening as human agreement increases.

4.2.3. Arm Count Comparisons

Finally, we compare our arm counts against those of the humans of GZ2 (Willett et al. 2013). During GZ2, when human classifiers indicated that a spiral arm pattern was visible, they were asked to determine the number of arms present in the galaxy using categories for one, two, three, four, or more than four arms, along with a "can't tell" option. These options are displayed in Figure 12.

Figure 13 assesses the extent to which humans agree with each other on arm count. We see that for some galaxies all humans agree on arm count, but in most cases human agreement is lower, with the maximum-vote arm count receiving only about 25% of the total human vote for some galaxies. This underscores the fact that arm count is ambiguous and disagreement among humans (i.e., between perceptions of what constitutes an arm) is common.

For each galaxy, we consider the arm count determined from GZ to be the category with the highest fraction of human votes.

Figure 12. Reference icons, from the GZ2 Web page, presented to human classifiers when asked how many spiral arms they saw.

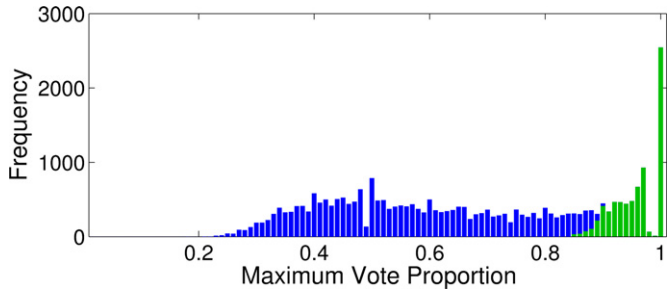

Figure 13. Humans find it difficult to agree on arm counts. For a given galaxy, the horizontal axis is the fraction of humans who voted for the most popular arm count for that galaxy, and the vertical axis is the number of galaxies (out of 29,250) that had the maximum vote. Maximum vote fractions are spread evenly from about 30% up to about 100%, indicating that humans often find it difficult to agree on the number of arms in a galaxy; a significant proportion of galaxies even have a maximum fraction less than 50%. The instances within the top human agreement quartile (as measured by Shannon entropy) are in green.

(A color version of this figure is available in the online journal.)

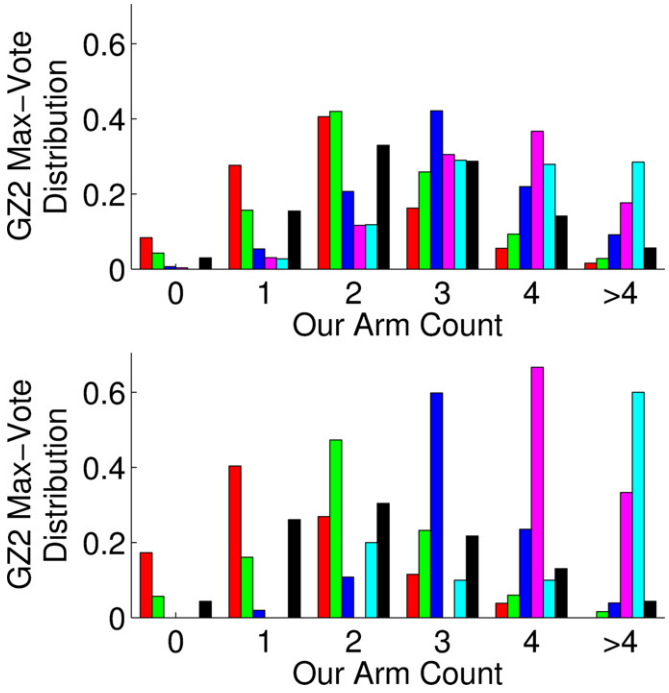

Figure 14. If we define an arm as any arc longer than 75 pixels, we get good agreement with humans on arm count. We plot distributions of the human maximum vote across our resulting arm count (zero, one, two, three, four, and more than four arms). The most popular arm count for humans is plotted by color, with red = 1, green = 2, blue = 3, magenta = 4, cyan = more than 4, and black = cannot tell. We use the maximum human vote because some galaxies do not have a majority human vote for any one category. In all arm counts produced from our output for comparison here, we only consider the arcs that agree with the dominant winding direction as determined by an arc-length-weighted vote. The top plot displays the distributions when all galaxies within our Galaxy Zoo comparison set are included. The bottom plot includes only galaxies within the top arm-count agreement quartile (lowest Shannon entropy quartile).

(A color version of this figure is available in the online journal.)

In some cases, no category received more than 50% of the vote, so we use the phrase maximum vote rather than majority vote. To produce arm count categories from our output, we count the number of arcs with length (in pixels) exceeding a threshold, including only arcs that agree with the dominant chirality; we have found that a threshold of 75 pixels gives relatively good agreement with GZ2 arm count, although we make no claim that this constitutes a sufficient or complete definition of an arm. For correspondence with GZ2 classifications, we group all of our counts above four within a single "more than four arms" category.

The GZ2 human vote count for two arms occurs much more frequently than the other categories, so we must ensure that indicators of performance on non-two-armed spirals are not overwhelmed by the relatively large number of two-armed spirals. To do this, we partition spiral galaxies by the most popular GZ2 arm count. For each group, we find the proportion of galaxies that our method assigns to each arm count category. These distributions are given in Figure 14. The top plot includes all 29,250 spiral galaxies, while the bottom plot uses only the galaxies in the highest quartile of human agreement (as measured by the Shannon entropy across human arm-count votes). When including all galaxies in the GZ comparison sample (top half of the figure), for some but not all "N arms" categories ($N \in \{1, 2, 3, 4, >4\}$), our method also counts N arms in the majority of cases (i.e., the distribution of galaxies with GZ2 maximum vote for N arms usually peaks at our measure of N arms). The bottom half of Figure 14 shows the same distribution across the galaxies in the top human agreement quartile. We now see that the human GZ2 vote distributions peak strongly at the matching SpArcFiRe arm count (for example, of the galaxies with a maximum human vote for three arms, in most cases SpArcFiRe also counts three arms, as shown in the distribution displayed in blue). Thus, our agreement with humans increases substantially when humans agree with each other.

5. DISCUSSION

We now discuss some of the insights gained while developing this software, especially ones that were not obvious at the start but turned out to be important.

First, the choice of representing a galaxy as a list of arcs (rather than using a more global model) was critical for representing general structure (although future work may be able to fit other models using our output). However, this choice does not come without cost. For example, if a grand design (i.e., symmetric two-arm) model encounters an image with two symmetric arms and a smaller possible third arm, the grand design model will ignore the arm that least conforms to the expected pattern. This would be the correct choice if the apparent third arm is noise, but our code considers the possibility that a true third arm exists. The result is that sometimes noise is

mistaken for a genuine arc, as can clearly be seen in some of the images in this paper. Sometimes such spurious arcs are short and can safely be ignored, but we occasionally find very long arcs that turn out to be spurious.

As has been the case in other computer vision tasks, we find that the choice of image features (information calculated from image brightness values) is important. In particular, when looking for galaxy structure as a list of arcs, gradient-based image information (and, more generally, local brightness patterns) can be more informative than the pixel brightness itself (although pixel brightness information can still be useful, and using it to a greater extent could be a subject of future work). This was recognized in Au (2006), where gradient information was represented as an orientation field. This orientation field is also applied in this work as information critical to our clustering procedure. Other gradient features could also be used; testing other gradient-based information is an area of further work that could potentially provide better results.

The requirement to center the image at the sub-pixel level (i.e., to precisely track the changes made to the fitted galaxy center during deprojection) had a surprisingly large impact. Because log-spiral arcs are defined in terms of the center of their coordinate system, even a sub-pixel shift in the estimated center of the galaxy can result in a slight change to some or all of the spiral arcs in the image. In the case of low-pitch-angle arcs, this change can spuriously flip the chirality of the arc. Such arcs can flip the measured chirality of the galaxy if they are long enough (individually or in aggregate). Even if such effects are rare, however, there may be a non-negligible amount of them in large image sets, so it can still be important to avoid them. We found that chirality flips in each direction were about equal, but in order to avoid the possibility of insidious biases, it is nonetheless desirable to avoid changes to the chirality (and all other measurements) when flipping the image, especially since potential biases have been important to consider in human classifications (see, for example, Land et al. 2008; Longo 2011).

When attempting to merge two fairly distant clusters that plausibly could be merged, it is problematic to merge them based on cluster shapes alone; classic hazards of extrapolation are difficult to overcome reliably. Over longer gaps, it becomes more problematic that arm segments are not always perfect log-spirals; when clusters are somewhat misaligned (in terms of the extensions of their arc fits), it is difficult to reliably distinguish between coincidental alignments and deviations from the log-spiral model. Additionally, large gaps often mean that the arcs to be merged are not very long, which increases fit error and uncertainty. Large gaps also require increased caution in arc merging (as such gaps are evidence, but not proof, that the arcs are not part of the same arm). Instead, to the extent that it is useful to merge clusters across long gaps, orientation information must be made more sensitive or other image features must also be used (along with the arc fits; which are obviously still useful). Proximity is one such image feature, but there are others that might also be helpful.

Interleaving local and nonlocal information is useful, as manifested through combining the pixel clustering with fit-based merge checking. The local information allows us to follow the observed shape of the spiral arms without relying too strongly on the log-spiral model, whereas the nonlocal information prevents clusters from encompassing more than one spiral arm (or an arm with two segments best modeled with separate arcs under the log-spiral model) where local information would otherwise suggest combining the two (e.g., at arm forks or bends).

Image brightness transformations can play an important role, as manifested through the use of the `asinh` function to convert from FITS to PNG-appropriate brightness, and by the use of the unsharp mask. When galaxy features are made more visible to humans, these features often become more visible to automated methods as well; making the arms more visible increases the proportion of the image brightness range covered by the spiral arms. However, it is challenging to apply a consistent brightness transform to a large set of images that may vary substantially in their brightness distributions; see, for example, Figure 5, where the very large but low-surface-brightness arm segment spanning the far right side of the image is entirely missed because it is too dim to be picked up by the orientation field. This is one place where experimentation with other image gradient descriptors may be useful.

More details and discussion can be found in Davis (2014).

6. FUTURE WORK

Referring back to Figure 4, we see an informal indication of how the parsing of an individual image can change depending upon minor changes to the algorithm. Additionally, given a fixed version of the algorithm, it is important to determine how our results change as a function of algorithmic parameters. We are in the process of performing a sensitivity analysis, as well as determining how our results degrade with degrading image quality. The results will be presented in an upcoming paper.

We have run our code on every object in the Sloan Digital Sky Survey (SDSS) that is 40 pixels across or larger and classified as a galaxy. Unfortunately, SDSS does not distinguish between spiral and non-spiral galaxies. We are currently working on an approach that uses the output of our code to distinguish between spiral and non-spiral galaxies. Because SpArcFiRe uses only shape and not color, we avoid problems such as accidentally misclassifying a red spiral or blue elliptical. Preliminary results are encouraging, and will be presented in a future paper. Once we can separate out spiral galaxies, further studies will be performed concerning how spiral structure correlates with other variables such as color, redshift, and local environment.

In the meantime, we have found informally that the existence of even one long arc is strongly suggestive of a spiral galaxy. Our database of arcs across the Sloan Survey can be queried to find objects satisfying certain quantitative shape criteria. For example, although our code does not explicitly search for rings, one can look for galaxies with one very long arm at a very low pitch angle. Figure 15 depicts nine typical objects found using this query. A forthcoming paper will present our code, as well as the results of running our code on every SDSS galaxy object in all available wavebands.

A Web interface to SpArcFiRe is currently available at http://sparcfire.ics.uci.edu. We will soon provide a distributable version of the code as well, both command-line driven and with a graphic user interface (GUI) front-end and back-end GUI arc viewer. Please contact W.H. (`whayes@uci.edu`) for questions, comments, or to request an early distribution.

We thank Steven Bamford for helpful insights, image sample selection, and prepublication access to Galaxy Zoo 2 classifications; Aaron Barth, Barry Madore, Scott Tremaine, Charless Fowlkes, Deva Ramanan, and Eric Mjolsness for helpful discussions; the anonymous referee for helpful suggestions; Sasha Volokh for downloading all the SDSS galaxy images; and the Arkansas Galaxy Evolution Survey (AGES) Collaboration for

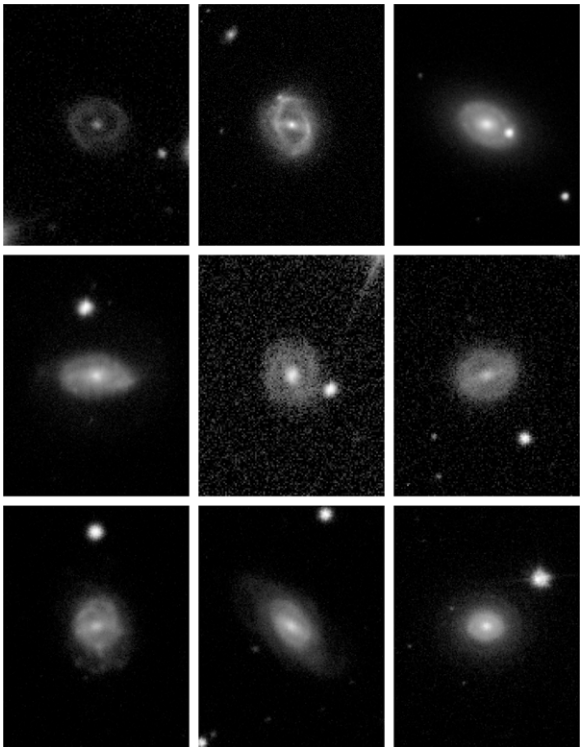

Figure 15. Some results of a query to our arc database: we look for galaxies with just one long, low-pitch-angle arm segment, yielding typical examples of ring galaxies. Most tend to have a bright compact object nearby that is likely the impacting object.

pitch angle measurements and discussions. Comparisons were also made possible due to image data from CGS as well as from SDSS and POSS II. Fellowship and travel support was provided by an ICS Dean's Fellowship at UC Irvine (for D.D.); the Oxford Centre for Collaborative Applied Mathematics; Steven Bamford and the MegaMorph project; and the AGES Collaboration (through NASA grant NNX08AW03A).

Funding for the Sloan Digital Sky Survey (SDSS) has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III Web site is http://www.sdss3.org/. SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration, including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofisica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University.

APPENDIX A

ALGORITHM CHANGES SINCE DAVIS & HAYES (2012)

Here we list significant changes that have occurred to the algorithm since Davis & Hayes (2012). These changes have significantly reduced the number of instances where our algorithm failed to produce output, and significantly improved our results when comparing against GZ humans (see our Table 6 versus Table 1 in Davis & Hayes 2012). Some of the items mentioned here are quite technical and would probably only be understood after reading Davis & Hayes (2012).

During ellipse fitting (see Figure 1), the center of the galaxy was previously estimated (tracked in deprojection) only to the nearest pixel. However, this is not accurate enough because in borderline cases, if we flip the image, then the arc fits can be changed slightly so that the chirality of each arc may not always flip in tandem with the image flip. In order to avoid potential biases we want the algorithm to always give exactly opposite chirality if the image is flipped. The problem was most prominent in low-resolution images. The galaxy center is now tracked more precisely, to the sub-pixel level, and now all but five flipped images (in our test set of 29,250) result in flipped chirality. (Note this is measuring self-consistency, not agreement with humans. We are 99.983% self-consistent in chirality, but agreement with humans is not quite so high.)

The clustering (see Figure 2(d)) was previously done in MATLAB but was moved to C++ for speed. Note that (for now) the main program is still in MATLAB, which calls C++ for clustering. This does not affect output per se, but it runs significantly faster, especially at higher resolutions.

Bright stars in the image can be a major distraction, and so must often be masked. In Davis & Hayes (2012), we used images given directly to us from the GZ team (S. Bamford, private communication) in which stars had already been masked. We now get all our images directly from SDSS, and perform our own star masking (although we can still accept images where stars have already been masked or otherwise removed, as was done both with our GZ1 comparisons and with the galaxies from the Arkansas group). If a bright star is present, then the ellipse fit depicted in Figure 1 sometimes attempts to include both the galaxy and the star (or in severe cases, only the star) into the interior of the ellipse. This causes the center of the ellipse to differ significantly from the center of the image. Because we assume that the galaxy is at least approximately centered in the image, this kind of star distraction can be detected automatically. We use SExtractor (Bertin & Arnouts 1996) to determine which parts of the image match the stars and then mask out those regions and retry the ellipse fit. This allows us to choose how aggressive of a star mask to use, when to apply it, and when not to apply it. Our algorithm now automatically progresses through several levels of star masking of increasing aggressiveness until the ellipse fit is within a certain distance of the center of the image; see Figure 16; even more detail is provided in Davis (2014). Even if we use a star mask during the ellipse fitting, we undo it after image standardization (see moving from Figure 1(e) to Figure 2(a)), so as to avoid unsharp-mask artifacts. This also avoids spurious arcs near star-masked regions.

Related to the above, we have made some changes to the brightness transformation from FITS to the PNG-scaled images we work with internally. In particular, we now do max-brightness clipping only after the brightness transformation; do brightness clipping using the max only within the galaxy region (as determined by SExtractor) rather than the whole image; and change brightness transformation quantile levels to bring the arms out more (reducing the number of instances where the standardization/ellipse-fit zooms in too far). Many of these measures simply compensate for the fact that we are now using our own star masks.

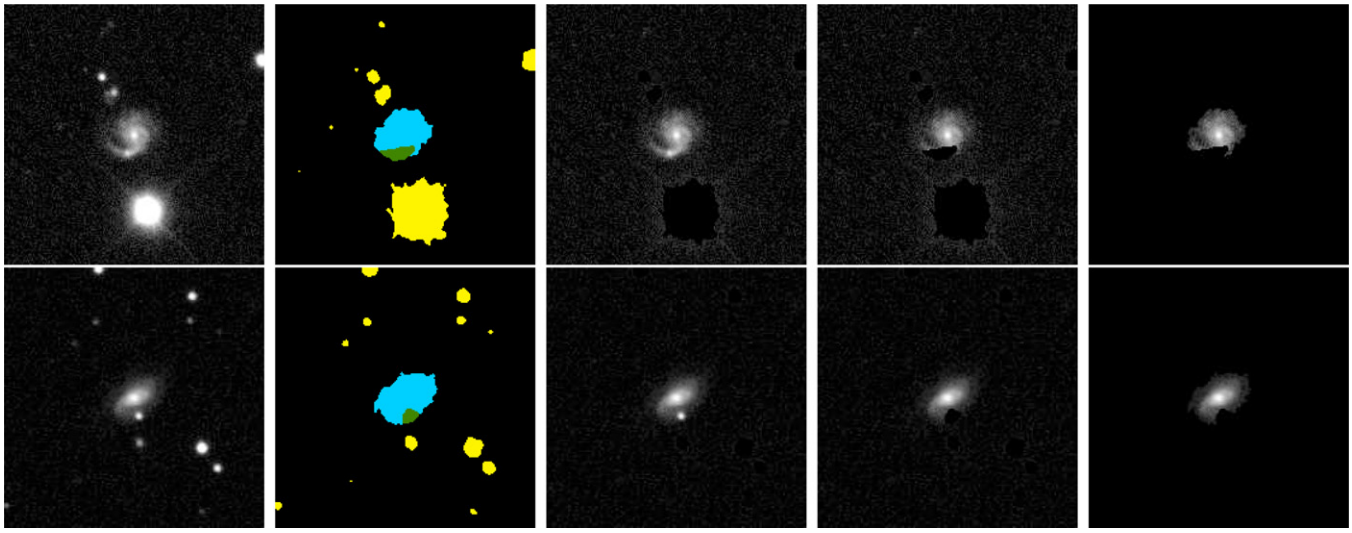

Figure 16. Star masking can prevent stars (and other objects in the image) from interfering with the image standardization depicted in Figure 1. The first column gives example images that contain objects other than the galaxy of interest. The second column shows image regions produced using cleaned SExtractor (Bertin & Arnouts 1996) output. The primary region is in blue, the part of the extended region that is not part of the primary region is in green, and SExtractor detections not part of either of these regions are in yellow. The next three columns show the different levels of star masking that can be applied using these regions. The first level (after the one that does not apply a star mask at all) removes SExtractor regions other than the primary or extended region (yellow regions). The second level also removes pixels that are in the extended region but not the primary region (green). Note that this removes a star in the top example and part of the galaxy in the bottom image, reflecting the fact that adjacent SExtractor detections are sometimes part of the galaxy and are sometimes a separate object. The final mask zeros all pixels except the ones in the primary region. This is done to accommodate the unlikely cases of missed SExtractor detections and bright pixels reassigned to background during SExtractor output cleaning. The last two masks sometimes zero out parts of the galactic disk, but even in these cases enough of the disk remains for the iterative Gaussian fit to remain close to the pixel region corresponding to the disk. After ellipse fitting the entire image is unmasked before processing continues.

(A color version of this figure is available in the online journal.)

The unsharp mask used for contrast enhancement can have edge effects near the boundary of the image, causing spurious orientation field vectors, and thus spurious arcs, along the edge of the image. These effects have been further minimized in the current code.

We have increased the sensitivity of the orientation filter, allowing us to reduce the strength of the unsharp mask.

We have also added a small diagonal to the covariance matrix during the ellipse fit to avoid numerical stability problems. This reduces the number of places where we cannot provide output. (The only remaining reason is that no arcs or clusters of sufficient size were detected in the image.)

As a supplement to bar detection, we now delete clusters that go through the center. Such center-containing clusters can be especially problematic when a highly inclined galaxy is deprojected, because the spherical central bulge of the galaxy becomes elliptical as a result of the deprojection, making it look a bit like a bar. This could possibly be fixed by performing bar detection entirely before image standardization, but will have to wait for future work.

Optionally, we can now apply a 3×3 median filter (to the original image, before image standardization) to reduce pixel-level noise.

APPENDIX B

MEAN DISCERNIBILITY AMONG GALAXIES IS THE SAME AS MEAN HUMAN CONSISTENCY

Assume we have a set of humans H who are classifying a set of galaxy images S. For simplicity we will assume that the classification consists only of choosing the chirality of a non-edge-on spiral galaxy, and that every human in H provides exactly one chirality choice for every galaxy in S. (In reality both S and H are huge and each galaxy in S is seen by only a few dozen members of H, but taking that into account would significantly complicate this short analysis and provide no better insight.)

Let g be a galaxy in S, and let D_g be the fraction of votes cast for the dominant chirality (S-wise or Z-wise). We call D_g the *discernibility* of g, and we note that $D_g \geqslant 1/2$ (because it is the dominant vote and there are only two choices in this simplified setting). Thus, D_g is both the fraction of humans that voted for the dominant chirality, and a measure of how clear that galaxy's chirality is to human observers.

Each human k casts one vote $v_{k,g}$ for each galaxy g, and we define that vote as

$$v_{k,g} = \begin{cases} 1, \text{ if } k \text{ voted for the dominant chirality of } g, \\ 0, \text{ otherwise.} \end{cases}$$

Note that the discernibility of g is thus

$$D_g = \frac{1}{|H|} \sum_{k \in H} v_{k,g}.$$

The mean discernibility of galaxies over the set S is

$$\bar{D} = \frac{1}{|S|} \sum_{g \in S} D_g.$$

Following Willett et al. (2013), we define the *consistency* C_k of human $k \in H$ over the set of galaxies $g \in S$ as the proportion of galaxies in S for which k's vote agrees with the dominant vote,

$$C_k = \frac{1}{|S|} \sum_{g \in S} v_{k,g}.$$

Consistency is meant to be a measure of how reliable each individual human k is. Some users are more reliable than others;

consistency provides a means to rank them, assuming that the group as a whole makes the correct choice.

Now, the mean consistency of humans over the set of galaxies S is

$$\begin{aligned}\bar{C} &= \frac{1}{|H|}\sum_{k\in H} C_k \\ &= \frac{1}{|H|}\sum_{k\in H}\frac{1}{|S|}\sum_{g\in S} v_{k,g} \\ &= \frac{1}{|H|}\frac{1}{|S|}\sum_{k\in H}\sum_{g\in S} v_{k,g} \\ &= \frac{1}{|H|}\frac{1}{|S|}\sum_{g\in S}\sum_{k\in H} v_{k,g} \\ &= \frac{1}{|S|}\sum_{g\in S}\frac{1}{|H|}\sum_{k\in H} v_{k,g} \\ &= \frac{1}{|S|}\sum_{g\in S} D_g \\ &\equiv \bar{D}.\end{aligned}$$

Thus, under the above assumptions, we see that the mean consistency among humans in H is equal to the mean discernibility of galaxies in S.

Refer now to the left half of Table 6, in which we do not insist our two longest arcs agree. We see that the bottom row suggests that our "length-weighted vote" measure of chirality provides a more consistent measure of chirality than the average individual human when the minimum discernibility is below about 95%; only when the chirality is completely obvious to humans (indicated by discernibility of 95% or higher) does our consistency drop slightly below that of the average human.[6] The same can be said for our measure "longest-arc alone," except the switch occurs at about 90% rather than 95%.

The right half of the table depicts cases where we are more confident of our choice because the two longest arcs agree. This measure applies only to a smaller set of galaxies, but in this set we see that our agreement with humans is even higher: both of our measures are more consistent than the average human until the humans get to virtually 100% certainty, where we still agree with the dominant choice in about 99.5% of these galaxies.

Ideally it would be nice to know the percentiles of human consistencies, and thus to rank SpArcFiRe's consistency among the humans of GZ, but to do so would require knowing the votes $v_{k,g}$ for individual humans, and this information is not public. Thus we can compare only to the mean human consistency rather than its median.

[6] Technically we have it easier than the humans: we are only making a binary choice; they had six choices.

REFERENCES

Au, K. 2006, PhD thesis, Carnegie Mellon Univ.
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Davis, B. L., Berrier, J. C., Shields, D. W., et al. 2012, ApJS, 199, 33
Davis, D., & Hayes, W. 2012, in 2012 IEEE Conference on Computer Vision and Pattern Recognition, (Piscataway, N.J.: IEEE), 1138
Davis, D. R. 2014, PhD thesis, University of California, Irvine
de Souza, R. E., Gadotti, D. A., & dos Anjos, S. 2004, ApJS, 153, 411
de Vaucouleurs, G. 1959, HDP, 53, 275
Duda, R. O., & Hart, P. E. 1972, CACM, 15, 11
Ho, L. C., Li, Z.-Y., Barth, A. J., Seigar, M. S., & Peng, C. Y. 2011, ApJS, 197, 21
Hubble, E. P. 1936, The Realm of the Nebulae (New Haven: Yale Univ. Press)
Land, K., Slosar, A., Lintott, C., et al. 2008, MNRAS, 388, 1686
Lintott, C., et al. 2011, MNRAS, 140, 166
Lintott, C. J., Schawinski, K., Slosar, A., et al. 2008, MNRAS, 389, 1179
Longo, M. J. 2011, PhLB, 699, 224
Ma, J. 2001, ChJAA, 1, 395
Nair, P. B., & Abraham, R. G. 2010, ApJS, 186, 427
Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, AJ, 124, 266
Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2010, AJ, 139, 2097
Perret, B., Mazet, V., Collet, C., Slezak, E., et al. 2009, in Image Analysis, (Berlin: Springer), 209
Ringermacher, H. I., & Mead, L. R. 2009, AJ, 137, 4716
Ripley, B. D., & Sutherland, A. I. 1990, RSPTA, 332, 477
Savchenko, S. S., & Reshetnikov, V. P. 2013, MNRAS, 436, 1074
Seigar, M. S. 2011, ISRAA, 2011, 725697
Seigar, M. S., Bullock, J. S., Barth, A. J., & Ho, L. C. 2006, ApJ, 645, 1012
Seigar, M. S., & James, P. A. 1998, MNRAS, 299, 685
Seigar, M. S., Kennefick, D., Kennefick, J., & Lacy, C. H. S. 2008, ApJL, 678, L93
Sérsic, J. L. 1963, BAAA, 6, 41
Shamir, L. 2011, ApJ, 736, 141
Simard, L. 1998, in ASP. Conf. Ser. 145, Astronomical Data Analysis Software and Systems VII, ed. R. Albrecht, R.N. Hook, & H.A. Bushouse (San Francisco, CA: ASP), 108
Willett, K. W., Lintott, C. J., Bamford, S. P., et al. 2013, MNRAS, 435, 2835