Title
Automated Quantification of Arbitrary Arm-Segment Structure in Spiral Galaxies

Permalink
https://escholarship.org/uc/item/0fv151p6

Author
Davis, Darren Robert

Publication Date
2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Automated Quantification of Arbitrary Arm-Segment Structure in Spiral Galaxies

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Computer Science

by

Darren Robert Davis

Dissertation Committee:
Professor Wayne Hayes, Chair
Professor Deva Ramanan
Professor Aaron Barth

2014

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am grateful to my advisor, Wayne Hayes, who supported me in myriad ways throughout my thesis research. Wayne suggested and created the opportunity to conduct research at this intersection of computer science and astronomy, provided guidance and perspectives critical to the development of this project while also enabling me to explore research directions that I found interesting, helped me navigate some of the complexities of graduate school, provided helpful comments on earlier versions of this thesis, and gave encouragement throughout my time at UCI.

I would also like to thank Aaron Barth for helpful insights about the capabilities, practices, and goals of observational astronomy; this perspective informed many stages of research planning. Aaron Barth and James Bullock gave important feedback about potential astronomical uses of my research.

Deva Ramanan, Charless Fowlkes, and Max Welling provided helpful guidance and a deeper understanding of computer vision and machine learning techniques.

Eric Mjolsness and Guy Yosiphon provided important early insights, including the state and potential of dynamical grammars for galaxy image parsing, and an improved understanding of the challenges facing efforts to extract information from spiral galaxy images.

Ben Davis, Daniel Kennefick, Marc Seigar, and others at the Arkansas Galaxy Evolution Survey (AGES) Collaboration provided a deeper understanding of the astronomical usefulness of pitch angle measures and their measurement through Fourier methods.

Steven Bamford helped select an appropriate image sample for comparison with human classifications in Galaxy Zoo 2, provided pre-publication access to Galaxy Zoo 2 classifications, and provided helpful insights.

Hundreds of thousands of volunteers classified images over the Web as part of the Galaxy Zoo project; their names are listed at `http://authors.galaxyzoo.org/`. These classifications enabled me to assess the accuracy of my method on large image sets. I would also like to thank the Galaxy Zoo team for building the platform enabling such classifications, and for establishing a phenomenon in citizen science that has inspired many to get involved or deepen their involvement in the scientific process.

Sasha Volokh downloaded the SDSS galaxy images needed for later comparisons with Galaxy Zoo and other work.

Barry Madore provided helpful suggestions about potential astronomical uses of my method. Roberto Abraham and Preethi Nair also provided helpful insights and discussions.

I would also like to thank my friends at UC Irvine for many interesting discussions and enjoyable experiences, and for helpful insights on research and graduate school. Lars Otten also shared LaTeX files that greatly facilitated dissertation formatting.

I am deeply grateful to my girlfriend, Lulu Zhang, for her support and understanding, especially during the demanding and time-intensive process of writing this thesis.

My parents have been extraordinarily dedicated and unwavering in supporting me throughout my thesis research and throughout my life. I am incredibly appreciative of all they have done in shaping and enabling who I have become.

My interest in astronomy began at a very young age, and became a lifelong interest in large part due to a “Striking Sparks” program that gives telescopes to young students enthusiastic about astronomy. In addition to my parents, who helped inspire and nurture my curiosity, I would like to thank my third grade teacher, Mrs. Chirco, for recognizing my interest, and for suggesting and supporting my entry into the Striking Sparks essay contest. I would also like to thank the Sonoma County Astronomical Society for organizing the Striking Sparks program and for building an inspiring environment of like-minded people who helped me develop my interest in astronomy and science.

I am grateful for the assistance provided by the Dean’s Fellowship at the Donald Bren School of Information and Computer Sciences, and for a summer 2011 visiting student fellowship from the Oxford Centre for Collaborative Applied Mathematics. The Donald Bren School of Information and Computer Sciences also provided conference travel support.

I have also had the opportunity to perform research in bioinformatics with Eric Mjolsness, and (separately) in bioinformatics and network analysis with Nataša Pržulj. I would like to thank Eric Mjolsness and Nataša Pržulj for helping broaden the scope of my research experience, and for the support provided by these research assistantships.

This project benefited greatly from the image data provided by several sky surveys, particularly the Sloan Digital Sky Survey (SDSS), the Carnegie-Irvine Galaxy Survey (CGS), and the Second Palomar Observatory Sky Survey (POSS-II).

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is http://www.sdss.org/.

The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Obser-

vatory, and the University of Washington.

The Digitized Sky Surveys were produced at the Space Telescope Science Institute under U.S. Government grant NAG W-2166. The images of these surveys are based on photographic data obtained using the Oschin Schmidt Telescope on Palomar Mountain and the UK Schmidt Telescope. The plates were processed into the present compressed digital form with the permission of these institutions.

The National Geographic Society - Palomar Observatory Sky Atlas (POSS-I) was made by the California Institute of Technology with grants from the National Geographic Society.

The Second Palomar Observatory Sky Survey (POSS-II) was made by the California Institute of Technology with funds from the National Science Foundation, the National Geographic Society, the Sloan Foundation, the Samuel Oschin Foundation, and the Eastman Kodak Corporation.

The Oschin Schmidt Telescope is operated by the California Institute of Technology and Palomar Observatory.

The UK Schmidt Telescope was operated by the Royal Observatory Edinburgh, with funding from the UK Science and Engineering Research Council (later the UK Particle Physics and Astronomy Research Council), until 1988 June, and thereafter by the Anglo-Australian Observatory. The blue plates of the southern Sky Atlas and its Equatorial Extension (together known as the SERC-J), as well as the Equatorial Red (ER), and the Second Epoch [red] Survey (SES) were all taken with the UK Schmidt.

All data are subject to the copyright given in the copyright summary (`http://stdatu.stsci.edu/dss/copyright.html`). Copyright information specific to individual plates is provided in the downloaded FITS headers.

Supplemental funding for sky-survey work at the ST ScI is provided by the European Southern Observatory.

CURRICULUM VITAE

Darren Robert Davis

EDUCATION

Doctor of Philosophy in Computer Science **2014**
University of California, Irvine *Irvine, California*

Master of Science in Computer Science **2012**
University of California, Irvine *Irvine, California*

Bachelor of Arts in Computer Science **2008**
University of San Diego *San Diego, California*

RESEARCH EXPERIENCE

Graduate Research Assistant **2009–2010, 2012–2014**
University of California, Irvine *Irvine, California*

TEACHING EXPERIENCE

Teaching Assistant **2010–2011**
University of California, Irvine *Irvine, California*

Grader **2005–2008**
University of San Diego *San Diego, California*

REFEREED JOURNAL PUBLICATIONS

SpArcFiRe: Scalable Automated Detection of Spiral Galaxy Arm Segments **2014**
The Astrophysical Journal (*Submitted*)

Revealing the Hidden Language of Complex Networks **2014**
Scientific Reports

REFEREED CONFERENCE PUBLICATIONS

Automated Quantitative Description of Spiral Galaxy Arm-Segment Structure **Jun 2012**
Computer Vision and Pattern Recognition (CVPR)

Exploring Content and Linkage Structures for Searching Relevant Web Pages **Aug 2007**
Advanced Data Mining and Applications (ADMA)

SEMINAR PRESENTATIONS

Measuring Structure from Spiral Galaxy Images **Nov 2013**
University of San Diego

ABSTRACT OF THE DISSERTATION

Automated Quantification of Arbitrary Arm-Segment Structure in Spiral Galaxies

By

Darren Robert Davis

Doctor of Philosophy in Computer Science

University of California, Irvine, 2014

Professor Wayne Hayes, Chair

This thesis describes a system that, given approximately-centered images of spiral galaxies, produces quantitative descriptions of spiral galaxy structure without the need for per-image human input. This structure information consists of a list of spiral arm segments, each associated with a fitted logarithmic spiral arc and a pixel region. This list-of-arcs representation allows description of arbitrary spiral galaxy structure: the arms do not need to be symmetric, may have forks or bends, and, more generally, may be arranged in any manner with a consistent spiral-pattern center (non-merging galaxies have a sufficiently well-defined center). Such flexibility is important in order to accommodate the myriad structure variations observed in spiral galaxies. From the arcs produced from our method it is possible to calculate measures of spiral galaxy structure such as winding direction, winding tightness, arm counts, asymmetry, or other values of interest (including user-defined measures). In addition to providing information about the spiral arm "skeleton" of each galaxy, our method can enable analyses of brightness within individual spiral arms, since we provide the pixel regions associated with each spiral arm segment. For winding direction, arm tightness, and arm count, comparable information is available (to various extents) from previous efforts; to the extent that such information is available, we find strong correspondence with our output. We also characterize the changes to (and invariances in) our output as a function of modifications to important algorithm parameters. By enabling generation of extensive data about spiral galaxy structure from large-scale sky surveys, our method will enable new discoveries and tests regarding the nature of galaxies and the universe, and will facilitate subsequent work to automatically fit detailed brightness models of spiral galaxies.

Chapter 1

Introduction

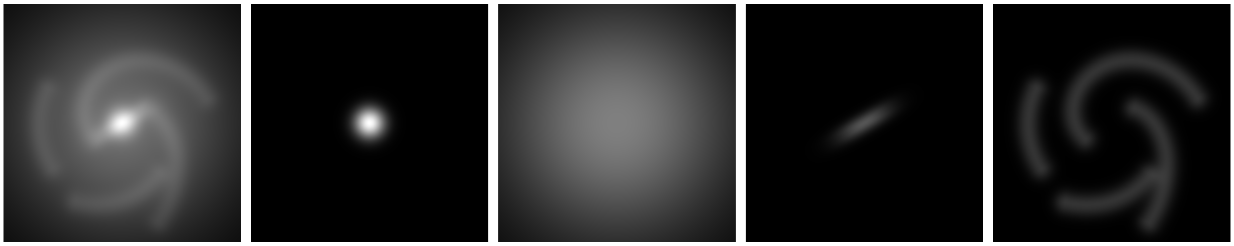
> "I often say that when you can measure what you are speaking about and express it in numbers you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science, whatever the matter may be."
> *Lord Kelvin (1824-1907)*

1.1 Background

Galaxies are extremely common phenomena in the known universe. Each galaxy consists of stars, gas, dust, and other material (likely including dark matter) held together by gravity. Galaxies have a wide variety of structures, but most have an elliptical or spiral pattern. Elliptical galaxies have a smooth, ellipsoidal shape with most of the brightness concentrated at the center; see Figure 1.1a for an example. Spiral galaxies, such as the ones shown in Figures 1.1b and 1.1c, have a visible bulge, disk, optional bar, and spiral arms. These components have a common center; for spiral arms this is the center of the spiral arm pattern rather than the spiral arms themselves. The bulge is a dense bright spheroidal region, the disk

(a) An elliptical galaxy (b) A spiral galaxy (c) A barred spiral galaxy

Figure 1.1: Example elliptical, spiral and barred spiral galaxies.



(a) Spiral galaxy (b) Bulge (c) Disk (d) Bar (e) Spiral arms

Figure 1.2: A schematic illustration of a spiral galaxy and its components, with a viewing angle orthogonal to the galaxy disk.

is a planar circular structure (unless severely gravitationally perturbed by another galaxy), and the bar has a linear pattern. The spiral arms can vary widely in arrangement, definition, and coherence, but almost always form a spiral pattern within the plane of the disk and around the galaxy center (non-planar arms could occur during a galaxy merger, but this is rare). These spiral galaxy components are illustrated in Figure 1.2. Lenticular galaxies have a bulge, disk, and optional bar, but lack spiral arms. Irregular galaxies do not have a consistent shape. Galaxies almost certainly also have a surrounding invisible dark matter halo with properties that can only be determined through the directly observable aspects of the galaxy.

1.2 Uses of Spiral Structure Information

Galaxy shape reflects forces at work within the galaxy, its neighborhood, and the universe as a whole, so galaxy structure reveals important information about the universe at each of these scales. As we will see in Chapter 2, automated methods are available for the precise determination of structure in elliptical and lenticular galaxies, as well as the bulge, disk, and bar components of spiral galaxies. Some basic information about the spiral arms (but not full descriptions of the spiral arm pattern in the general case) can also be obtained automatically with existing methods. Chapter 2 also reviews human-guided study of detailed and general spiral arm structure in small samples of spiral galaxies, as well as manual classifications of structure in large samples of spiral galaxies. The continued existence of these labor-intensive manual efforts strongly suggests that spiral galaxy structure information is valuable for learning about galaxies and the universe.

At the scale of the universe as a whole, galaxy measurements (including spiral galaxy measurements) can be used in order to test several hypotheses about cosmological evolution. One of these is the cosmological principle, i.e., the assumption that at sufficiently large scales, the universe appears the same regardless of the position or viewing direction of an observer. Tests of this principle have been conducted using information about galaxy positions (and other data, such as measurements of the cosmic microwave background). There is also interest in using spiral galaxy structure information, particularly the distribution of spiral galaxy winding directions, to test other implications of the cosmological principle. If the cosmological principle holds, galaxies should not have a preferred winding direction as a function of location on the night sky. Contrary to this expectation, Longo [41] found that among about 2,800 spiral galaxies selected by eye for spiral pattern clarity and winding direction, there was a preferred spin direction along an axis. Land et al. [34] found a similar pattern in human classifications from the Galaxy Zoo project [40]. However, by presenting the human classifiers with mirrored and unmirrored images (and with other analyses), Land et al. also found that the human classifiers tended to identify S-wise spiral patterns more readily than Z-wise spiral patterns (possible explanations include higher sensitivity to S-wise patterns in the human visual system or biases induced by the design of the Galaxy Zoo website). Since the sample of galaxies in the Longo study (and the Land et al. study) was skewed much more heavily toward galaxies in the northern hemisphere of the sky, Land et al. found that this preference for S-wise galaxies could explain the apparent alignment in spiral galaxy spin. Later, Longo [42] repeated his study with more (about 15,000) spiral galaxy images and random mirroring in the images displayed to the human classifiers (without any indication of whether mirroring was used for a particular image), finding an even stronger signal for preferred spin direction along an axis. Using an automated tool to determine winding direction, Shamir [63] also found preferred spin direction along an axis (with a slightly different axis that remained within error bounds) using an even larger sample. However, winding directions were only obtained for about 36% of the galaxies in the sample. Additional unbiased (and likely automated) measures of spiral galaxy winding direction will thus be useful to fill in these gaps and provide further tests as to whether spiral galaxy winding directions are in line with the cosmological principle.

Winding direction is perhaps the simplest measure of spiral galaxy structure, and so winding direction is just the beginning of what spiral galaxy structure can reveal about galaxies and the universe. Despite commonalities in possessing a spiral pattern, spiral galaxies exhibit many variations in arm arrangement, including arm tightness (and the distribution of arm tightness), the presence and type of asymmetry, the number, position, and angle of arm forks, and the overall arm count, with many of these factors having nuances of their own. Each of these properties could give clues about the evolution of spiral galaxies and the universe as a whole, since we can measure these structural parameters as a function of galaxy position, and of the age of the universe when the observed galaxy light was produced. Additionally, by investigating associations between spiral galaxy structure and environment (e.g., how many galaxies are nearby), it will be possible to learn more about how galaxy structure changes due to interactions with other galaxies.

Spiral arm structure is also critical to understanding galaxy dynamics because it provides information about the dynamical states of galaxies, and has observed and/or predicted correlations with other spiral galaxy properties. For example, the density wave theory of spiral structure [36] predicts that spiral arm tightness correlates with the galaxy's central mass concentration. Additionally, spiral arm tightness has been observed to correlate with the mass of the black hole at the galaxy core [6] and properties of a galaxy's dark matter halo [60]. As we will see in later chapters, the automated methods described in this work can provide data about spiral galaxy structure that was previously unavailable for large samples of galaxies. This detailed large-scale spiral galaxy structure information will provide the data needed to look for and evaluate possible new correlations, which can in turn provide insights about the interplay of different aspects of spiral galaxies. Where measurements of spiral galaxy structure already exist, alternate modes of measurement would be useful in order to provide independent corroboration or, if the measures conflict, to prompt re-examination. This assessment of measurement confidence is especially important because some galaxy properties were determined subjectively or semi-subjectively. Automated spiral galaxy structure measurement can also test whether correlations and other findings of previous quantitative studies extend to larger samples (before the availability of the method described in this work and in [12, 13], detailed quantitative structure was only known for a few tens of galaxies, but we have run our method on the hundreds of thousands of Sloan objects classified as galaxies, and can readily run our code on results from future surveys). These samples could not only have greater statistical power, but could also have improved generality because larger samples can often reach a greater portion of the observable universe (which, even if the cosmological principle holds, is only homogeneous at a sufficiently large scale). Where new and previously-discovered correlations are found to extend to these larger samples, automated description of visual structure could be used as proxy estimates of properties that would otherwise need to be examined with a process more expensive or otherwise less available than imaging (such as spectroscopy).

With automated methods for determining spiral galaxy structure, it would also be possible to search a large set of galaxies for objects where quantitative measures suggest worthiness for in-depth human study. For example, astronomers may wish to study galaxies with certain structural properties, such as rings. By running automated structure-description methods on a large set of images, the resulting database could be queried in order to produce a list of objects of interest for the study. For example, to find galaxies with rings, one could look for cases where there is a long arm with very low pitch angle. It can also be useful to identify "weird" (qualitatively distinct) galaxies for further study. In the Galaxy Zoo project, spontaneously identified unusual objects have turned out to be interesting (e.g., [39], [9]), and automated methods would be a useful complement because they would likely identify different objects as unusual. Automated methods would not have the human intuition for weirdness but could, for example, identify subtle but unusual deviations from a typical pattern. Quantitative measures suggesting spiral galaxy peculiarity could include extreme structure measurements or galaxies that do not have a typical relationship between properties. These measures could be useful on their own, or spiral structure information could provide additional signals to peculiar-galaxy detection methods such as the one in [62], potentially making these methods more sensitive to abnormalities in spiral structure.

At the scales of the universe, galaxy clusters, and individual galaxies, automated determination of spiral galaxy structure will provide measurements crucial to testing theories about galaxies and the universe. In support of goals such as these, large scale sky surveys such as the Sloan Digital Sky Survey [69] (SDSS) have been conducted, with more planned or in progress. SDSS has images of almost a million galaxies, many of them spirals. Despite the large scale of SDSS, orders of magnitude more galaxy images will be available with future sky surveys due to increased coverage and sensitivity. The Hubble Ultra Deep Field [5] found about 10,000 galaxies in 1/13,000,000th of the sky, implying that there are about 100 billion (i.e., 10^{11}) galaxies observable from orbit at the sensitivity and wavelengths used. With the sensitivity of the Hubble Extreme Deep field [30], the estimated galaxy count is closer to 200 billion. Even more galaxies are likely to be observable at even higher sensitivities and at other wavelengths (especially the infrared), including the sensitivities and wavelengths of the planned James Webb Space Telescope.

For studies involving spiral galaxy structure, the enormous quantities of galaxies imaged in large-scale sky surveys need to be translated from pixels to structure information. With this structure information, it will be possible to compare theoretical predictions to observed structure, and to assess the extent to which simulations match observations of the universe (with some of these simulations producing images that also need to be analyzed for spiral galaxy structure). These capabilities are fundamental to the scientific process as applied to galaxies. Thus, large scale spiral structure information promises to reveal more about galaxies (including fundamental matters such as the origin of spiral arms, which still remains at least partially unresolved) and about the universe itself. Large-scale sky surveys provide the imaging data, but manual efforts to translate it into spiral galaxy structure information will become increasingly difficult as the number of observed galaxies continues to increase.

1.3 Primary Contributions

After reviewing related work in Chapter 2, I will describe a method for determining quantitative spiral galaxy structure automatically. As discussed in detail in Section 3.1, galaxy structure will be described as a list of logarithmic spiral arcs, as one would obtain by individually tracing arcs along each spiral arm (using an arc with a well-defined mathematical shape). This spiral arm “skeleton” is general enough to capture the wide variety of arm arrangements and other structure variations found in spiral galaxies. As shown in Figure 1.3, this is critical for accurate description. When simpler measures (such as winding direction and average arm tightness) are desired, they can be calculated from the spiral arm skeleton information. Galaxy brightness is not explicitly included in this representation, but pixel regions corresponding to each arc are also reported, reducing the problem to a brightness fit of a known image region with a single galaxy component of known functional form. In Section 2.2, we will see that automated methods are already available for brightness fitting in similar situations.

To the best of my knowledge, this method is the first to determine general spiral galaxy structure in a fully automatic manner. This gives precise, quantitative information, without human perceptual bias and with full repeatability. The method described in this work is intended to allow extraction of spiral galaxy structure data useful for the astronomical purposes discussed in Section 1.2, to facilitate automated brightness profile fitting of spiral galaxies with tools currently requiring extensive human input (e.g., GALFIT [51]), and to provide insight into the use of image processing and image understanding techniques in astronomical applications.

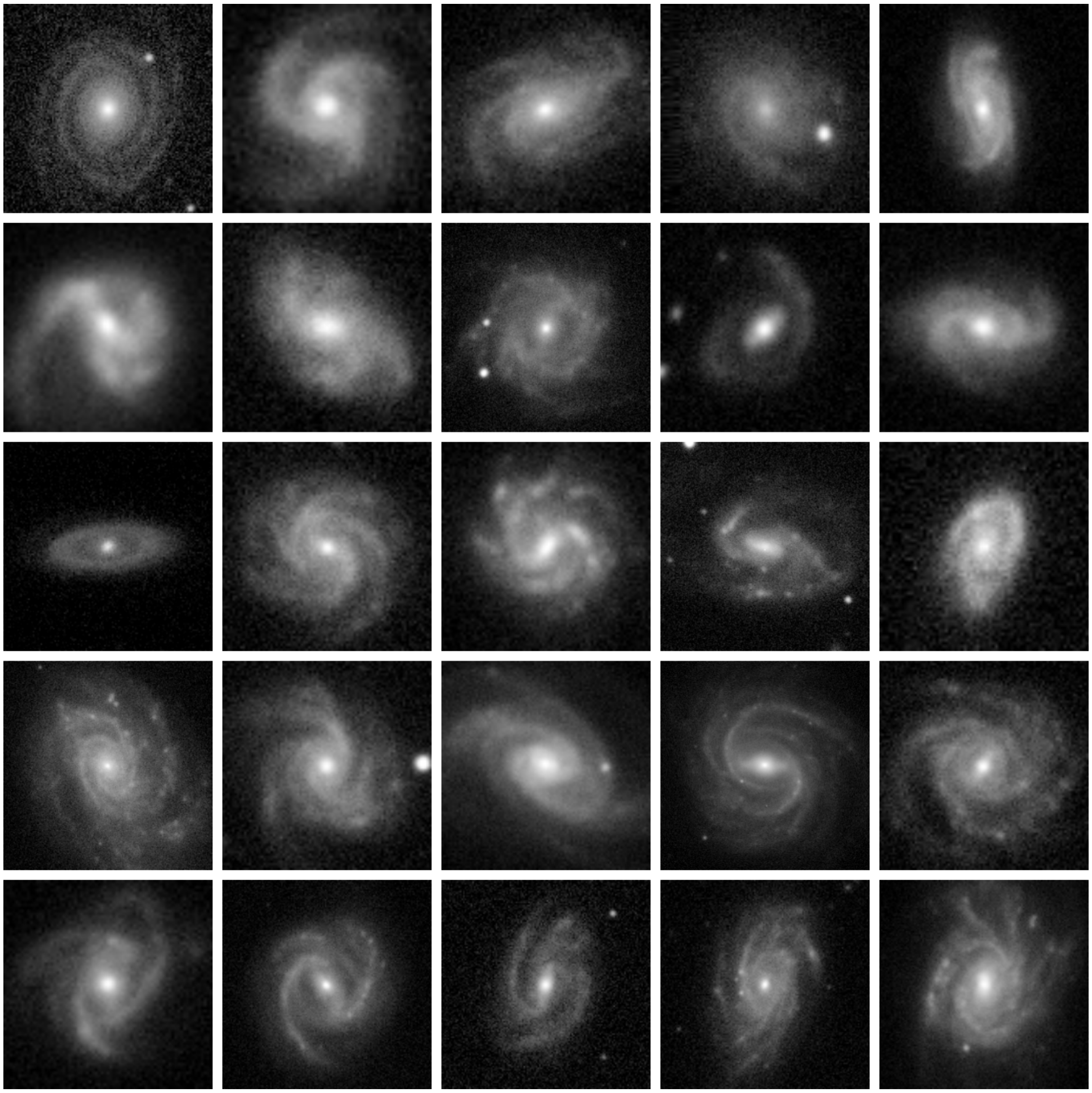


Figure 1.3: Spiral galaxy structure descriptions must be very flexible in order to capture the myriad variations observed for this type of galaxy. Spiral galaxies can vary in arm count (a measure that is often not even well defined), may not be symmetric, can have arm forks and arms that are not spatially contiguous with other arms or other galaxy components, can have severe bends in the arms, and, in general, can have almost arbitrary arm arrangement.

Chapter 2

Related Work

2.1 Morphological Classification

Galaxy classification systems are one of the earliest schemes for characterizing galaxy structure. The Hubble sequence [29] was the first classification system to gain widespread popularity, and it is still in use today. The categories are typically arranged in a "tuning fork" pattern (Figure 2.1), starting with elliptical galaxies, which are categorized according to their apparent degree of ellipticity. Next is a category for lenticular galaxies, which have a bulge and flat disk, but no visible spiral arms. Spiral galaxies are split into two sequences of categories according to whether the galaxy has a bar or not. Within each of these sequences, the category is chosen according to the prominence of the galaxy bulge (particularly the concentration of light near the center of the galaxy), how tightly the spiral arms wind around the center of the galaxy, and the extent to which the spiral arms are smooth versus fragmented. Several extensions of this categorization system have been proposed; for example, the de Vaucouleurs system [16] uses three parallel sequences for spiral galaxies (barred, unbarred, and intermediate), adds a category at the ends of these sequences for galaxies with fragmented arms and a weak bulge, and also introduces a three-way branch for the presence of rings (ring structure present, absent, or intermediate). For a review of galaxy classification systems, see [66].

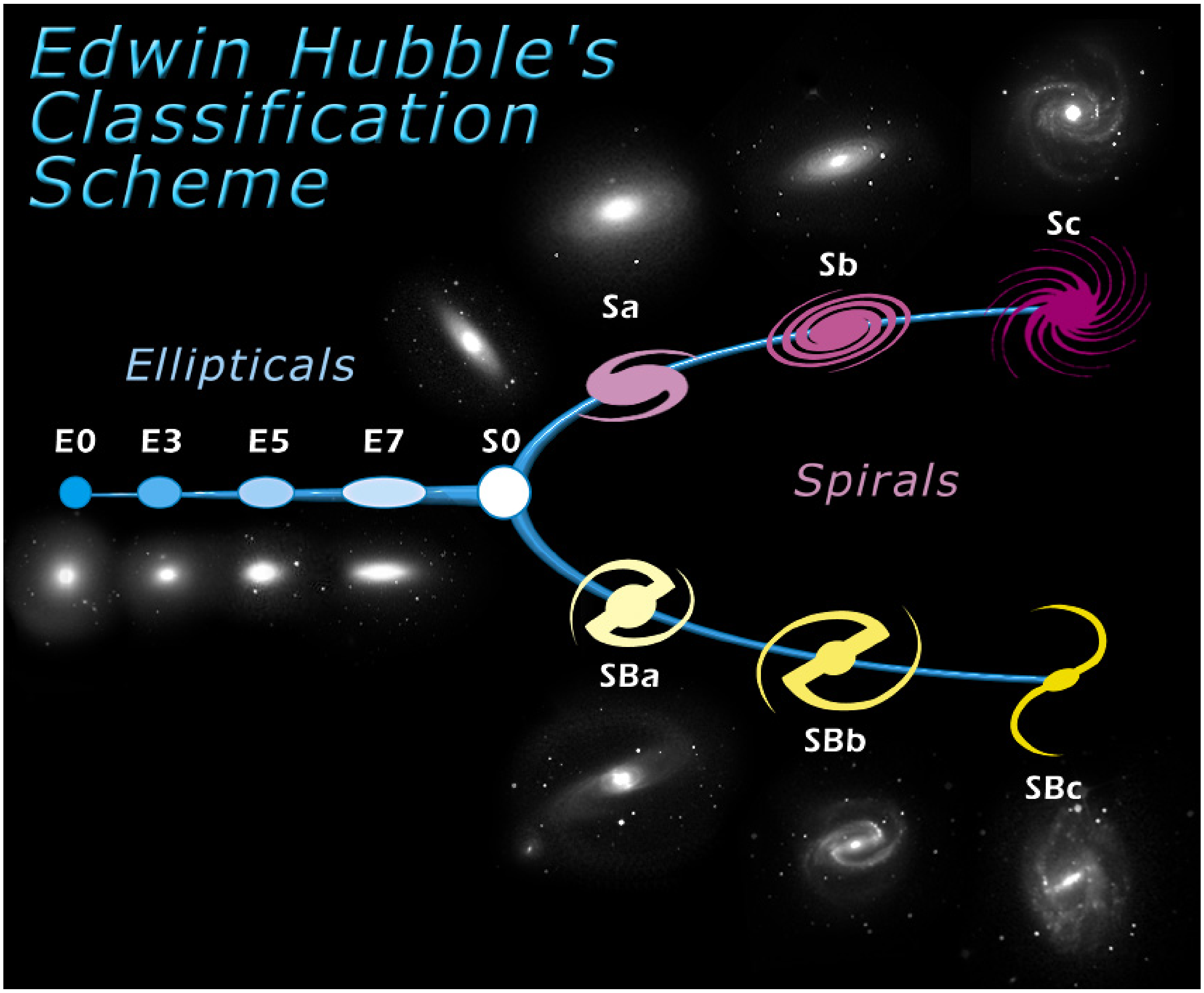


Figure 2.1: The Hubble “tuning fork” classification system. The fork begins with elliptical galaxies classified according to their degree of apparent ellipticity, followed by S0 (lenticular) galaxies that have a bulge and disk, but no visible arms. The branches correspond to unbarred and barred spiral galaxies, with categories along this sequence arranged according to the prominence of the galaxy bulge, as well as the tightness and smoothness of the spiral arms. Irregular galaxies, which do not have well-defined shapes as elliptical, lenticular, or spiral galaxies, are not shown. Image credit: STScI/NASA.

Category membership gives useful information about the structure of spiral galaxies, and correlates with important physical properties [55]. Consequently, many automated methods describe spiral galaxy structure by determining a galaxy’s category membership within one of these classification systems (see [4] for a review of these methods). However, the Hubble sequence and related classifications are inherently subjective [55], in part because classification criteria are not precise and may conflict with each other (for example, a galaxy may have tightly-wound arms, suggesting it should be placed toward the middle of the fork in Figure 2.1, but also have a weak bulge, suggesting it should be placed toward the end of the fork). Human classifiers thus use their own criteria to remove ambiguities and resolve conflicts. Automated classifiers trained on these categorizations simply attempt to recreate the interpretations used by human classifiers. Even if conflicts between categorization criteria were minimized (e.g., by using an automated procedure that is allowed to define the categories, and optimizes the distinctness and consistency of these categories), the categories themselves discard useful structure information, in part by discretizing quantitative measures of spiral arms (e.g., their winding tightness). Furthermore, since the categorizations are done at the galaxy level, they tend to ignore or make assumptions about spiral arm arrangement (e.g., asymmetry and arm forks) and ignore within-galaxy variation in arm properties. This information loss could in theory be addressed with a very large number of categories, but this would be impractical due to the granularity needed to capture quantitative measures, and the exponential growth in categories for each combination of galaxy-level criteria and for each spiral arm. Thus, galaxy classifications are useful, but for full information it is necessary to determine quantitative properties of individual galaxy components (including the bulge, disk, bar, and spiral arms).

2.2 Bulge, Disk, and Bar Fitting

Automated methods are available to fit quantitative models of the galaxy bulge and disk (and, in some cases, the bar as well). Compared to spiral arm patterns, these components are relatively simple. Each type of component appears at most once in the galaxy, and they always have the same center (or nearly the same center) as the galaxy itself. All of these components also have brightness contours well represented by a generalized ellipse (an ellipse where the exponent is also a parameter, allowing the ellipse to vary in shape from a 4-tipped star to a rectangle). Thus, each component can be described by a two-dimensional brightness model with parameters for the generalized ellipse that determines the brightness contours, and for the brightness fall-off as a function of the distance defined by these contours.

Since the bulge, disk, and bar can each be described with a fixed number of parameters, typical data fitting techniques can be adapted for astronomical use. In some works, the bulge and disk are fit by directly using a χ^2 fitting routine or a nonlinear least squares solver. Several software packages are also available for performing these bulge and disk fits (as the primary feature or as one of their capabilities), including GIM2D [64], BUDDA [15], GALFIT [50], and MegaMorph [24]. BUDDA, GALFIT, and MegaMorph are also able to fit bars. Fitting the bulge, disk, and bar gives precise information about these components, and can make remaining features (including spiral arms) more visible by subtracting other light. However, they do not fit the spiral arms themselves.

2.3 Spiral Arm Pitch Angle Measurement

Since spiral arm structure has many types of variation not encountered with the bulge, disk, and bar, automated fitting is much more difficult in this case. However, if the spiral arms are assumed to be symmetric (which disregards many types of potential variation including some of the cases illustrated in Figure 1.3, but nonetheless is appropriate for some galaxies), then a measure of spiral arm winding tightness (the pitch angle) can be determined quantitatively. This can be done with a Fourier transform that represents the image as a sum of logarithmic spirals (or other spiral functions) [56]. In its basic form, this procedure entails correcting for galaxy inclination (so that the image appears as if the galaxy were viewed along a line orthogonal to its disk), choosing inner and outer radii (the radii of the annular ring that defines the pixel region used in the Fourier transform), finding the Fourier component (logarithmic spiral function) with the highest amplitude as a function of the pitch angle and the number of Fourier modes (corresponding to the number of arms when spiral structure is highly symmetric), and then using the pitch angle of this Fourier component. In [11], additional human supervision is added for reliability: the pitch angle is plotted as a function of the inner radius to determine inner-radius ranges where the pitch angle is stable, and solutions (Fourier components with a given pitch angle and mode count) are selected according to visually determined agreement with the image and visually determined stability in the pitch angle plots. As described in [57] and [59], Fourier methods can also be used to measure pitch angle as a function of distance from the galaxy center. This is done by computing a series of Fourier transforms, each covering a smaller, contiguous portion of the galaxy distance range, and varying the inner and outer boundaries of this ring-shaped window. The galaxy distance range (i.e., the smallest inner-radius boundary and the largest outer-radius boundary) is still determined manually.

These Fourier methods provide important quantitative information with a well-defined mathematical foundation. However, they assume symmetry in spiral arm structure and they require manual input to determine the Fourier window and to interpret the results. The symmetry assumption is reasonable in some cases, but many galaxies are strongly asymmetric, so Fourier models are not always appropriate. If different arms in the galaxy have different pitch angles, or if at least one arm does not cover the full radial extent of the galaxy, pitch angle measures (overall or as a function of radius) may be misleading. In the common case of flocculent spirals, which consist of a larger number of scattered arm fragments, asymmetry can be especially severe, and the lack of spatial localization in the Fourier transform could make it difficult to distinguish these shorter, weaker arms from coincidental brightness alignments. Additionally, manual input is needed in order to determine the inner and outer radii for the ring-shaped Fourier window; the inner radius is especially important because excessively large values miss some of the spiral structure and excessively small values introduce severe interference from the galaxy bulge (and sometimes other components, such as the bar). Human supervision is also useful in order to verify that the chosen Fourier component is a good visual match with the image. In many settings, these factors are not limiting; many galaxies are symmetric, and manual supervision is common and appropriate in many astronomical studies. However, other methods are needed for strongly asymmetric galaxies, for image sets too large for per-image human input, and when arm-level details about spiral structure are desired.

2.4 Automated Fitting of Restricted Spiral Models

Automated methods for spiral galaxy fitting are available for restricted classes of spiral galaxy models. In [2], the spiral model consists of two symmetric spiral arms attached to the two ends of a (possibly zero-length) bar. Orientation (i.e., the strength and direction of local linear structure) is determined for each pixel, and then the model is fit by minimizing the difference in alignment between the model and the local orientation near the model. Coarse-to-fine grid search is used to determine the model parameters because gradient descent-like methods encountered too many local optima. The symmetric, two-armed spiral model is also used in [52], but the fit uses brightness information instead of local orientation, and more detail is added, namely parameters for the bulge and disk, parameters for the brightness distribution of each component, modeling of the patchiness from HII regions seen at shorter wavelengths, and support for multiple wavebands (but with only the brightness parameters, not the shape parameters, variable across wavebands). Error bars for the estimated parameters are also given. The model is fit with a Markov chain Monte Carlo procedure with simulated annealing, using different annealing temperatures for different parameters, and fitting the model in three main stages (first the bulge and disk, then with the bar added, and finally the full model). Both of these methods can be useful for precise study of potentially large samples of "grand-design" spiral galaxies (which have two smooth, prominent, symmetric spiral arms), but many galaxies do not follow this pattern.

Dynamical grammar models [45] are applied in [70] in order to represent spiral galaxy structure as a set of stars (or "on" pixels) belonging to knots that in turn arise from spiral arms. To fit this model, the number of arms and knots must be pre-determined, and then an expectation-maximization algorithm [17] is used to determine the assignment of stars to knots, the assignment of knots to arms, and the (logarithmic-spiral-like) shape parameters for the spiral arms. Results are demonstrated for a simulated and a thresholded image. The dynamical grammar model is promising for representing arbitrary spiral galaxy structure (and even the processes underlying this structure), but the fitting is complex (and can reach local optima) even for a model with the arm count fixed to two. By using automated methods that can determine the position and approximate shape parameters of the spiral arms (i.e., the methods described in this work, specifically Chapter 3), it may be feasible to automatically fit a dynamical grammar model that expresses the full range of spiral galaxy structure.

Ganalyzer [61] is a tool that assesses the degree to which a galaxy is spiral or elliptical. Ganalyzer works by correcting for galaxy viewing angle, producing a two dimensional radial intensity plot that gives image brightness as a function of distance (from the galaxy center) and polar angle, applying peak detection (as a function of angle) for each fixed discretized distance value, grouping the peaks, and then fitting lines (in the space defined by the radial intensity plot) to the top one or two groups of peaks to determine the extent to which the galaxy is spiral vs. elliptical (since spiral galaxies are expected to have stronger shifts in their brightness peaks as a function of distance from the galaxy center). In addition to determining the spiral-ness of a galaxy, Ganalyzer's intermediate representation is likely sufficient to provide a measure of spiral arm tightness, and in [63] it was used to measure spiral arm winding direction. Ganalyzer is very fast and promises to be useful for obtaining some galaxy-level spirality measurements, but does not provide a model of the spiral galaxy structure itself, and we are not aware of any attempts to evaluate accuracy beyond spiral/elliptical classification for 525 galaxies [61] and spiral arm winding direction for 120 galaxies [63].

2.5 Human-Interactive Detailed Spiral Fitting

Other methods for spiral structure extraction use human guidance as direct input to the spiral fitting procedure. This human interaction has been necessary in order to determine spiral galaxy structure without sacrificing generality. An early work in this area is [54], which represents spiral arms by a series of connected line segments parameterized by their starting position and the angles between adjacent line segments. Given a good, manually determined initial positioning of these spiral arms (automatic determination of these initial positions was attempted but not successful), the fit is refined using a Markov chain Monte Carlo technique.

To determine pitch angle variations along spiral arms (in the galaxy M51) without the symmetry assumptions of Fourier analysis, [48] examines wavelet coefficients as a function of image position and wavelet orientation. Starting with a manually selected point on each spiral arm, wavelet-coefficient maxima are used to trace spiral arms (arm spurs and forks are deliberately ignored), and the maximum orientation at each such point gives the local pitch angle. This has enabled detailed study of arm positions and pitch angles for a single, manually examined galaxy with prominent spiral structure.

In [44], points are manually selected at the start of and along each spiral arm, and then a logarithmic spiral arc is fit to the points selected for each arm. A similar procedure is followed in [14], where human insight about spiral dust lanes or bright stars is incorporated into the point selections, in [58], where this manual selection of points is one of the two methods used to measure pitch angle (the other is a Fourier method of the type discussed in Section 2.3), and in [27] to study the spiral structure of the whirlpool galaxy (M51). In [26], the start location of each spiral arm is manually determined, and then the pitch angle (and galaxy inclination) are manually adjusted to obtain a good fit. These methods are suitable for detailed study of individual galaxies or small sets of galaxies, and allow incorporation of expert knowledge about the galaxy or about the task at hand (for example, if one wants to trace star-forming regions of an arm, those points can be selected). Such methods can also describe arbitrary spiral structure because the points on a spiral arm can be freely selected without requiring arm symmetry or other assumptions. However, this manual selection introduces subjectivity, reduces repeatability, and limits the number of galaxies that can be practically examined.

For detailed fits of galaxy brightness distributions, GALFIT [50, 51] is typically used. GALFIT allows specification of an arbitrary number of brightness components and their initial (pre-fit) parameters. In earlier versions of GALFIT [50], these brightness components had generalized-ellipse contours, allowing fits to galaxy components such as bulges, disks, and bars using one or more brightness components for each of these galaxy components (GALFIT is flexibile enough to accommodate user judgment as to how many brightness components are used, and the initial parameters for each one). The brightness fall-off function and its initial parameters are also chosen for each component. The most recent version of GALFIT [51] can model spiral structure (and other non-axisymmetric galaxy features) by allowing modifications of the contours of each component's generalized ellipse. The contours' distance from the center (as a function of polar angle) can be modified by sinusoidal and/or polynomial functions, and spiral rotations can be introduced into the coordinate system used as the frame of reference for the contours.

Since GALFIT allows an arbitrary number of galaxy brightness components, each with a wide array of available shape parameters, GALFIT can fit arbitrary spiral structure. Each of these components, as well as good initial values of all free parameters and the selection of which parameters are fixed, must be provided by the user for each galaxy. Without per-galaxy specification and unless the model is simple enough to characterize structure that is fairly similar across all galaxies (such as the bulge and disk), GALFIT's minimization procedure (a Levenberg-Marquardt algorithm using the χ^2 model-to-image distance) will encounter many local optima or other problematic conditions, and fail to reach a satisfactory fit (though GALFIT is generally able to detect that this has happened). As such, GALFIT is well suited to provide precise refinements of galaxy models manually developed for each galaxy, and has seen wide success in detailed study of relatively small sets of galaxies, but it is not intended or suitable for automated fitting of spiral galaxies. However, the methods described in Chapter 3 may provide the means to automatically determine initial per-galaxy spiral models and parameters suitable for input to GALFIT.

2.6 Large-Scale Manual Determination of Spiral Galaxy Structure

To determine spiral galaxy structure over large sets of images, fully manual efforts are still used. The Galaxy Zoo project [40] coordinates hundreds of thousands of human volunteers to classify Sloan Digital Sky Survey [69] images by eye over the Web. In the initial Galaxy Zoo project, volunteers were asked to place images into one of six categories: clockwise or anticlockwise spiral, edge-on spiral, elliptical, star/don't know, or merger [40]. Galaxy Zoo 2 [67] adds increased detail; for spiral galaxies this includes categories for the number and tightness of spiral arms, the relative dominance of the bulge, and the presence or absence of a bar. Elsewhere, a painstakingly detailed structural catalog of 14,034 galaxies was manually derived by a single professional astronomer [46]. Designations include galaxy T-type (morphological classification under the de Vaucouleurs system, as discussed in Section 2.1) as well as features such as bars, rings, and gravitational lensing, and spiral arm counts and flocculence. Human classifications of spiral arm winding direction have also been conducted for hundreds to thousands of galaxies on multiple occasions ([65], [8], [68], [31], [33], [1], [42]); in particular, 15,158 galaxies were classified by eye in [42]. These manual determinations of spiral galaxy structure are useful in part because they can incorporate human judgment and common sense, but human perceptual biases are a concern, quantitative information is difficult to obtain, and manual input will be increasingly difficult as image set sizes continue to increase.

2.7 An Unfulfilled Need for Automated Fitting of General Spiral Structure

We have surveyed a variety of methods for galaxy structure determination. Each type of method has its own mixture of structure-extraction abilities and limitations, and there are situations where each type of method is useful. However, there are also situations where one would like to automatically extract general spiral galaxy structure, and this is not covered by existing methods: one has been required to either restrict the amount of spiral structure information available (e.g., by using classifications, galaxy-level measurements, or a non-general model of spiral galaxy structure), or eschew automation. Our method, described in Chapter 3, will fill this gap by giving a description of the spiral arm "skeleton" as a list of parameterized arcs that can describe arbitrary spiral arm structure. This method does not give information about the brightness distribution of the galaxy or its components, but our method could be combined with GALFIT [51] or other brightness-fitting methods in order to do so. In addition, since the method provides the pixel regions corresponding to each spiral arm segment, these pixel regions could be used to determine brightness-related measures of individual arm segments.

Chapter 3

Extracting Spiral Arm-Segment Structure

> "I never worry that all hell will break loose. My concern is that only part of hell will break loose and be much harder to detect."
>
> *George Carlin (1937-2008)*

3.1 Representing Arm-Segment Structure

In order to extract information about spiral galaxy structure, it is first necessary to have a sense of what a spiral arm is. Astronomers have studied spiral arms in detail, but to the best of my awareness, a precise and broadly adopted definition of "spiral arm" does not exist in the astronomical literature. Instead, spiral arms tend to be an "I know it when I see it" phenomenon: astronomers tend to represent the concept of spiral arms though examples and visual description rather than through a precise definition.[1] Perhaps this is because spiral

---

[1] This is not without precedent; consider the long period of time in which the term "planet" was used before being formally defined, as well as the controversy surrounding this definition, despite extensive prior study of the various objects frequently referred to as "planets."

arms are common, visually distinctive, and subject to precise study on the level of individual galaxies, but exhibit a wide range of variation among galaxies.

Galaxies often have two arms, but many others have one, three, four, or even more arms. Spiral arms may be coherent for a long span, or they can exhibit various degrees of *flocculence*, with highly fragmented spiral arm structure. They may or may not be symmetric, may or may not have forks, may or may not be connected to structures such as a bar or the bulge, and, in general, can have almost arbitrary arrangement. See Figure 1.3 for a depiction of some of the variation observed in spiral arm structure.

Most of these types of variation concern the arrangement of spiral arms, rather than the arms themselves. An important consequence is that individual arm shapes are consistent enough to allow at least approximate description with a parametric function. In 1942, Danver [10] proposed using the logarithmic spiral arc shape, noting it was a good fit for observed galaxies, and it has been commonly used since then. Other functions have been considered, but the logarithmic spiral is perhaps the most widely used, and is regarded as providing a good fit to spiral arms (see, for example, [11] and [14]). Since spiral arms can be at least approximately described using a logarithmic spiral arc shape, since (as discussed later) we can accommodate deviations from the spiral arc shape (such as arm bends) by using more than one arc and by reporting the corresponding image region along with the arc, and since the logarithmic spiral function is already common in astronomy, we use this function to describe individual spiral arms (or segments of arms). Logarithmic spirals can be expressed in polar coordinates as

$$\mathrm{lgsp}_{\phi,a,r_0}(\theta) = r_0 \cdot e^{-a\cdot(\theta-\phi)}, \quad 0 \leq \theta - \phi \leq \theta_e \tag{3.1}$$

Here, ϕ gives the angle between the polar axis and the starting point of the arc. r_0 is the initial radius, which defines the distance from this starting point to the origin, and thus

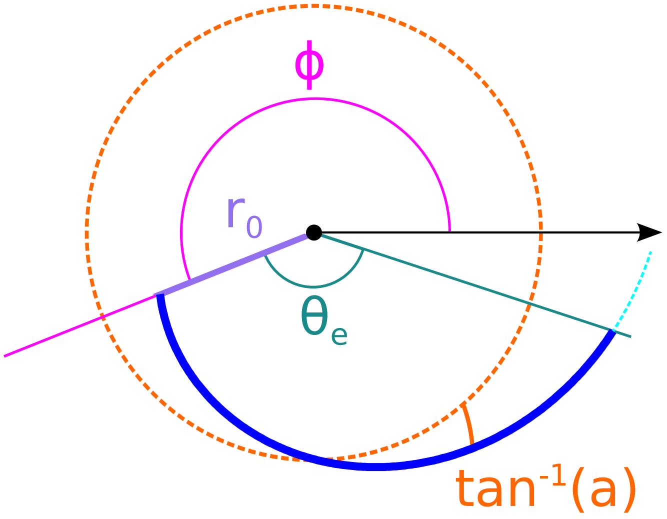


Figure 3.1: Parameterization of the logarithmic spiral in Equation 3.1. The dark blue region denotes the arc given by the parameters, and the light blue region shows a partial continuation of the arc beyond its angular extent θ_e. The initial radius r_0 controls the size of the arc and ϕ rotates the arc. The most important parameter is the pitch angle; taking the arctangent of this value gives the angle between the arc and a circle centered at the origin. This angle is constant throughout the span of the arc.

affects the size scaling of the arc. a is the pitch angle; taking the arctangent of this value gives the angle between the arc and a circle centered at the origin. This angle is constant along the spiral curve. A positive pitch angle means that the arc winds inward from its starting point, and a negative pitch angle means that the arc winds outward. θ_e specifies the angle subtended by the arc (this angle may exceed 360 degrees). ϕ and θ_e then define the arc's endpoints; the arc starts at $\theta = \phi$ and ends at $\theta = \phi + \theta_e$. These parameters are illustrated in Figure 3.1.

An automated system should be able to determine these spiral arm shape parameters, while also accounting for the many types of variations seen in galaxy-level spiral structure (particularly in arm arrangement). As such, the choice of structure representation is critical. If it is not general enough, it forces information loss and biases results toward structure that can be represented. Since we aim to produce an automated method, we must also ensure that the structure representation does not demand information that is only available through human input, and remains computationally practical on large sets of galaxy images.

As discussed in Section 2.1, galaxy classifications (e.g., the Hubble type) give a rough indication of spiral structure but lose information due to the differences in structure that can be found within each category. Another choice has been to use a galaxy-level model with a fixed number of parameters – for example, a model with two symmetric arms extending from a possibly zero-length bar, with parameters to control the bar length, arm start location (bar rotation), arm tightness, and possibly the position and inclination of the galaxy as a whole. While such models can describe certain subtypes of galaxies (e.g., "grand design" spirals with exactly two arms that are long, well-defined, and symmetric), they cannot cover the full range of variation in spiral galaxy structure.

It is then natural to consider using a mixture model, i.e., to choose or automatically infer a set of galaxy models (i.e., mixture model components) and have the fitting procedure select a model (or weighted set of models) that best matches an input image. However, mixture models assume an underlying classification scheme (where the classifications are the choice of or weights on mixture model components), but spiral galaxy structure does not fall neatly into categories. For example, one could consider defining mixture model components according to the number of spiral arms in the galaxy, but spiral arms vary widely in how well defined they are, to the point where arm count can be ambiguous. Even if the number of arms were known, variations in their arrangement would either force a large number of sub-categories (such as how many arms are forks of another arm, connected to a bar or the bulge, or freestanding), or require within-category variation to be addressed through means other than mixture models. Although these factors do not make mixture modeling impossible, they do suggest that mixture modeling is not a natural choice for representing spiral galaxy structure.

Part-based models [20] can alleviate some of the rigidity in shape models. Instead of using a single global template, part-based models represent visual phenomena as a set of parts along with probability distributions on the spatial arrangements of pairs of parts. This decomposition into parts and their spatial relationships is appealing, but part-based models assume that the objects of interest can be represented with a tree structure that is consistent across cases. Visual variations (e.g. missing parts) can be accommodated, but are treated as deviations from a "true" structure (e.g., some parts may not be visible because they are occluded by some other object). Part-based models have been successful in areas where the consistent tree structure assumption holds (such as person detection [19]; the set of human body parts is roughly constant at the needed level of detail, and these parts are attached to each other in a consistent way). However, the previously-discussed variations in spiral galaxy structure entail that galaxies do not have a fixed tree structure. Consequently, the use of part-based models would still require a choice between the loss of generality in single models and the explosion of component-model possibilities in mixture models. Furthermore, a prominent advantage of part-based models is that they can encode information about which spatial arrangements of parts are more likely than others, but due to the current lack of large-scale quantitative data about spiral galaxy structure, favoring certain types of spatial arrangements would only bias the results. Even if this data were available, spiral galaxy structure is likely too inconsistent for spatial-arrangement preferences to be reliable, although spatial arrangement signals (e.g., arm endpoints are somewhat more likely near the galaxy bulge) could eventually be a useful refinement as long as they are weak enough to avoid biasing the results.

Other models would be able to represent the full variety of spiral structure, but contain so much complexity themselves that it becomes difficult to determine which model variant best matches the image. As discussed in Section 2.5, GALFIT II models [51] can express a wide variety of spiral structure through arbitrary sets of component parts that can be individually positioned. The structure expressed in these models would be extremely valuable, but such models pose difficulties for automatic fitting, even with the large degree of expert human input needed in order to select components and determine reasonable initial parameters for these components. As such, these models currently have practical limits on the structure that can be extracted automatically, even if they can represent arbitrary structure in great detail. However, once spiral structure is available automatically through other representations, the extra information may enable automated fitting of GALFIT models by guiding a search through the highly complex model space that the GALFIT models impose. Visual grammars [72] likely also have the requisite flexibility to represent spiral arm structure. However, visual grammars are so general that they would be a setting for, rather than a full answer to, questions of spiral galaxy structure representation. Visual grammars would also introduce challenges of their own: a model of sufficient complexity may pose issues with computational tractability, and grammar rules could produce biases that favor certain spiral arm structures over others. Existing (and likely simpler) spiral galaxy structure information would likely be needed to guide the search for model configurations and parameters for each galaxy, so that computational tractability can be retained. Such information is also needed in order to validate that the model covers all variations in spiral galaxy structure, and does not prefer certain structure configurations over others unless the favored configurations are proportionally more common. Consequently, it is prudent to pursue automatic fitting of other, simpler representations of spiral galaxy structure before attempting to fit a visual grammar model. Dynamical grammars [45] can capture even more information, since grammar rules can represent physical processes that give rise to spiral arm structure, rather than just describing the visual aspects of this structure. Such models are very promising because they can directly encode hypotheses and models of galaxy dynamics, along with their relationship to the observed image. In [70], a two-armed spiral galaxy grammar model was fit to a few galaxies (see Section 2.4), but dynamical grammar models used for fitting can be much more detailed and general once automated methods can extract prior information (about general per-galaxy spiral structure) that identifies promising areas of the (very large) dynamical grammar search space. Like GALFIT and visual grammars, dynamical grammars are likely to be more appropriate for automated fitting once more information about spiral galaxy structure is available to validate models and guide fitting. Dynamical grammars would then be able to refine and expand upon this information.

In consideration of the need to represent a wide variety of spiral structure while also retaining computational practicality across large datasets, I have chosen to represent spiral galaxy structure as a set of logarithmic spiral arcs (Equation 3.1) in a shared polar coordinate system, along with the pixel regions corresponding to each arc. This allows precise quantification of individual spiral arms, whose "backbone" structure is typically close to a well-defined parametric arc shape (the logarithmic spiral). Crucially, this precision does not come at the expense of generality. Since the logarithmic spiral arcs are individually positioned, this representation can naturally accommodate wide variations in arm arrangement, and thus in galaxy-level spiral structure. At the level of individual arms, most arms follow the logarithmic spiral shape, but even when a single logarithmic spiral cannot describe the entire arm with suitable accuracy, a sequence of a few such arcs can be placed along the spiral arm, improving generality and thus accuracy, with the additional benefit of revealing unusually-shaped arms. The logarithmic spiral function and coordinate system are two dimensional because spiral arms are roughly coplanar (they reside within the galaxy disk). Using a shared coordinate system forces the spiral arms to have a common center, but this does not lose useful generality because the existence of a common center is a known physical property of spiral galaxies.

A key difference between this list-of-arcs representation and the other representations discussed above is that the arms are modeled (and can be detected) separately, which bypasses complications that would otherwise arise from structure variations. To model a fork in an arm, for example, two arcs can be arranged to produce this fork shape. We do not need to explicitly create a mixture model component, grammar rule, or other model construct that accommodates this situation, do not need to ensure that such model constructs explicitly cover all structure variations (requiring manual or automated examination of a large set of spiral galaxy images), and do not need to expend computational time explicitly searching for forks and other structure variations. Instead, both parts of the fork are arcs that can be detected individually. The fork can be found afterward by examining the arrangement of the arcs (e.g., by considering the closest distance between endpoints of one arc with any point on another arc). Additionally, since the set of arcs can have arbitrary size, this representation does not create restrictions or rapid search-space expansion with respect to arm count.

The list-of-arcs representation does not take advantage of possible patterns in relationships between components. However, at this point this would be more of a liability than an advantage because the patterns would be very loose, we don't have data to infer these patterns even if they exist, we would have to make sure to avoid biases, and there would be additional computational overhead. This representation also does not model brightness information, but this facilitates independence between the spiral arcs, avoiding the detailed parameter interactions that are one reason why GALFIT models are difficult to fit (see [51] and [49] for further discussion of fitting considerations). Furthermore, the list-of-arcs representation provides the information needed to determine brightness parameters with existing methods. The pixel region and logarithmic spiral shape would be known, so it would only be necessary to calculate the brightness fall-off as a function of distance from and along the logarithmic spiral arc (or the arc parameters can be refined as part of the fit, but they still do not need to be estimated from scratch), possibly in combination with bulge and disk fits to the galaxy. As discussed in Section 2.2, fitting brightness models with a fixed number of parameters can already be done automatically.

In addition to the arc-level parameters, it is also helpful to include some information at the galaxy level, namely the galaxy center and inclination (viewing angle). Since material in all spiral arms orbits a common center, and all spiral arms are coplanar, the center and inclination are common across all spiral arms in a galaxy, and affect all of a galaxy's spiral arms in a consistent manner. Thus, determining the center and inclination at the galaxy level avoids a problematic choice between re-discovering this information for each spiral arm and entangling the detections of the individual spiral arms (which would lose the important benefits of separate detection discussed above). Using these galaxy-level parameters also improves the speed and reliability of the spiral arm fits (since there are fewer free parameters), and guarantees consistency of information that is known in advance to be consistent. As will be discussed in Section 3.3, the center and inclination can be determined before starting detection of spiral arms, so this variation can be removed before spiral arm fitting. Furthermore, since the center and inclination are consistent within any particular galaxy, fixing these parameters within a galaxy does not lose any useful generality.

Since the list-of-arcs representation produces well-defined arcs, but the notion of a spiral arm is not well defined in all cases, I will refer to the former as an arc, will refer to the region of the arm described by the arc as a "spiral arm segment," and will reserve the term "spiral arm" for the (sometimes ambiguous) reference to the physical phenomenon in spiral galaxies. Arcs should, of course, describe spiral arms, but there may be more than one arc (and thus more than one spiral arm segment) per spiral arm. This can happen when it is not obvious whether one spiral arm is a continuation of another. For example, if a spiral arm has a tight bend, it may be better (for semantic and/or modeling purposes) to describe it with two spiral arm segments, one on each side of the bend, and leave it up to the user (and the task at hand) to decide whether the spiral arm segments belong to the same spiral arm. Other ambiguities include brightness gaps that could result from dust lanes or two distinct arms, and branch (fork) structures where either (or neither) of the two prongs could be considered the continuation of the spiral arm. Additionally, the option to use multiple arm segments to describe (what could be interpreted as) one spiral arm allows additional modeling flexibility; if a spiral arm shape deviates too much from a logarithmic spiral curve, the arm can be described with multiple arm segments instead of a single arm segment. In light of these considerations, 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.

The list-of-arcs representation describes a spiral arm-segment “skeleton” that provides crucial information about spiral galaxy structure. From the set of arcs it is possible to determine quantities such as arm winding direction, tightness (pitch angle), asymmetry and lopsidedness, branching, and degree of flocculence. This information will be useful for the astronomical goals described in Section 1.2. It can also facilitate development of further automated methods, which can use this information as initial conditions or constraints on the fit. Furthermore, to the best of my knowledge, this representation allows the most detailed description of spiral arm structure currently available from an automated method. In the following sections, I will describe a method for determining the information given under this representation.

3.2 Image Brightness Transformation

The method described in this work can accept images in popular formats (e.g., PNG or JPEG), in which case the steps described in this section can be skipped, but astronomical images are generally acquired and used as FITS files. These files contain close-to-raw data from a telescope’s sensor (i.e., as flux measurements), and as such these files differ from most images familiar in everyday life. Of particular importance is that FITS images have a very high dynamic range: some image pixels have brightnesses orders of magnitude greater than other pixels, depending on where they were pointed (much more brightness can be received from a galactic bulge as compared to a region without a visible object, for example). This poses extra hazards when applying image processing and computer vision methods to these images (high-brightness pixels can have outlier-like effects on linear filtering methods, for example). In addition, brightness values from spiral arms encompass only a small portion of the image's brightness range, so without further processing, large brightness values from sources like foreground stars and galactic bulges would drown out the signal (brightness variation) from spiral arms. To address this, we transform image brightness values so that spiral arm-related variation covers a much wider portion of the image intensity range.

Astronomers already have methods for brightness transformation, in part because this operation is also useful when preparing images for human viewing. We adopt the procedure described in [43], where brightness values at each pixel are transformed according to the function

$$r(x) = \begin{cases} 0 & \text{if } x < m \\ \dfrac{\operatorname{asinh}(\frac{x-m}{\beta})}{\operatorname{asinh}(\frac{M-m}{\beta})} & \text{if } m \le x \le M \\ 1 & \text{if } x > M \end{cases} \tag{3.2}$$

Astronomers frequently use the asinh function because, as discussed in [43], it is approximately linear for smaller brightness values and approximately logarithmic for larger values. The linear region of the function preserves details from relatively faint regions, such as spiral arms. The logarithmic part suppresses extreme intensities from regions that are much brighter than spiral arms, preventing these regions from dominating the intensity range of the image. The parameter β controls the brightness level where the function transitions from linear to logarithmic. The behavior of this function is illustrated in Figure 3.2.

We must now choose values for m, M, and β. When choosing m, we note that the smallest brightness value in the image is generally significantly larger than zero, due to near-constant values for the background sky. If we were to set $m = 0$, $r(x)$ would compress all of the brightness values toward the upper end of the transformed brightness range, making the entire image appear bright and inducing severe information loss. A reasonable response is

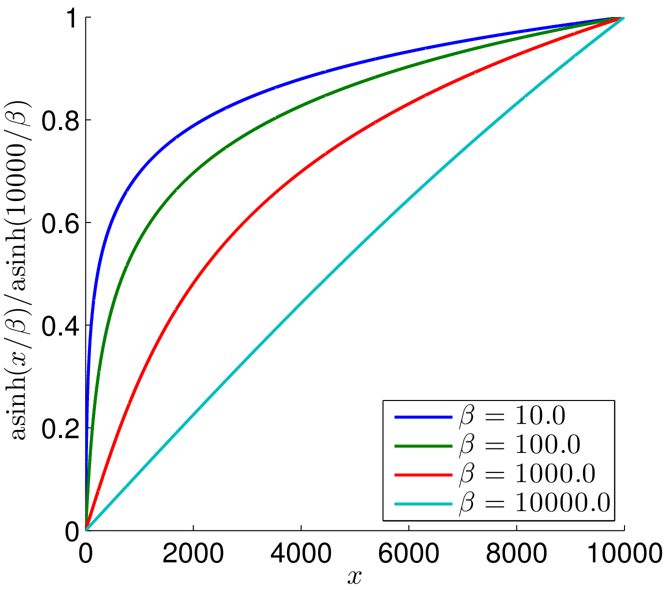


Figure 3.2: Behavior of the asinh function and its β parameter. The function is applied after subtracting a near-minimum brightness value from all pixels in the image (see text). This asinh function is approximately linear for $x < \beta$, retaining brightness variation from features such as spiral arms, and approximately logarithmic for $x > \beta$, suppressing extreme brightness values. Smaller values of x (corresponding to relatively faint features such as spiral arms) thus span a much larger range of the output values. Appropriate settings for β (discussed in the text) can thus transform image brightness values in a way that emphasizes spiral arms.

to set m to the smallest observed pixel value. This can be improved because empty (for our purposes) regions of the image vary in brightness, in significant part due to noise from the image sensor. We do not want these noise values to be prominent in the brightness range of the image, especially since $r(x)$ retains the most brightness variation at the lower end. Consequently, we set m to the 25^{th} percentile of image brightness values. This cuts off much of the brightness variation from image noise, adds robustness against unusually small pixel brightness values, and remains conservative enough to preserve faint objects in the image. Setting the value for M is simple: since the asinh function suppresses extreme brightness values, we do not need a maximum-brightness cutoff. Consequently, we can simply set M to the maximum brightness value in the image.[2]

---

[2] The M parameter is only present because [43] uses the more generic $F(x)$ instead of $\text{asinh}(x/\beta)$, and then suggests $F(x) = \text{asinh}(x/\beta)$. With the more general $F(x)$, which we do not need here, the M parameter may be useful.

The parameter β is the most important because it determines the transition point between linear and logarithmic brightness transformation. Its value need not be exact, since brightness variation from the spiral arms need not occupy an exact portion of the image's brightness range, but it must still be set to a reasonable value. If it is too high, the brightness range will still be dominated by bright, non-arm regions. If it is too low, brightness variation from the arms themselves will be suppressed, and they will be difficult to distinguish from the rest of the galaxy. We cannot use a fixed value for β because image brightnesses vary according to factors such as the range of light wavelengths recorded in the image, the instruments used, and (in some cases) the exposure time. Since extreme-brightness regions occupy only a small portion of the image area, we can use a brightness percentile to find a good value for β. In particular, we set β to the 75^{th} percentile of the image brightness.

Since β must be chosen before the brightness transformation is performed, the β value selection does not benefit from the increased consistency imparted by the brightness transformation. Additionally, a single asinh operation is sometimes insufficient to fully emphasize spiral arms over extreme brightness values. These shortcomings are removed by applying $r(x)$ twice. After the first transformation by $r(x)$, we treat the resulting image as input, repeat the percentile-based selection of m and β (using the same percentiles as before), and apply the second asinh transformation.

Figure 3.3 illustrates the effect of the brightness transformation. The original image appears mostly black, with only a few small bright areas visible. Since most of the brightness variations are miniscule in comparison to the difference between typical and extreme values, the very brightest image regions dominate at the expense of all other regions. In the transformed image, the brightness range of the spiral arms (and other galactic features) is preserved, while the difference between typical and extreme brightnesses shrinks dramatically. To both the human eye and to automated methods, the brightness information in the image now focuses much more on features of the galactic disk, including spiral arms.

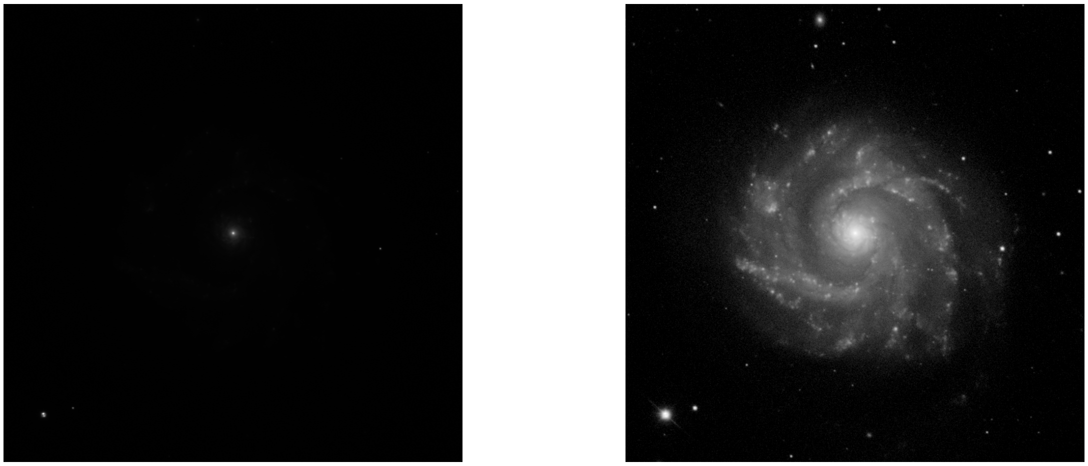


Figure 3.3: A sample galaxy image (NGC3631, red band, from the Sloan Digital Sky Survey) before (left) and after (right) brightness transformation. Without transforming the image brightness, the image appears mostly black because the bright objects (namely the galactic bulge and foreground stars) are much brighter than everything else in the image, including the spiral arms. After transforming the image brightness, intensity variations from the spiral arms occupy a much wider range of the image values, making them much more visible to the human eye and to automated analysis procedures.

3.3 Image Standardization

Since spiral arm structure is roughly two-dimensional (the spiral arms lie within the plane of the galactic disk), it is ideal to view a galaxy *face-on*, such that the image plane aligns with the plane of the galaxy disk (i.e., where our line of sight is orthogonal to the galaxy disk). In a few cases, galaxies are indeed viewed face-on, but it is more common for the galaxy to be positioned so that the disk is tilted (inclined) relative to our line of sight. If this situation is left unaddressed, the two-dimensional geometry of the galaxy disk will not correspond to the geometry in the image. In particular, the spiral arms will appear distorted with respect to a logarithmic spiral shape. Fortunately, unless the inclination is so severe that spiral arm structure is not visible at all (an *edge-on* view), it is possible to compensate for inclination effects. Along with inclination, variations due to the galaxy center and apparent size will also be detected and removed.

After determining the galaxy inclination, center, and apparent size, the image will be transformed so that the galaxy appears as if it were viewed face-on (except for the spherical bulge, which will appear elliptical for severely inclined galaxies), centered in the image and at a fixed resolution. It is also possible to leave the image as-is and adjust all subsequent steps for galaxy inclination, position, and size, but transforming the image ensures that all steps adjust for these factors in a consistent manner. Transforming the image also avoids adding unnecessary algorithmic complications and computation time to all subsequent steps.

The transformation to a standardized inclination and center makes these quantities fixed and known, thus providing the galaxy-level information needed to use the list-of-arcs representation of spiral structure (see Section 3.1). Standardizing the resolution of the image makes size-related galaxy properties (e.g. the number of pixels covered by a spiral arm, or the length of an arc describing an arm) consistent between galaxies. This aids the orientation filter approach described in Section 3.5 (since there is less relative variation between spiral arm sizes and orientation filter sizes), increases the consistency of size-related decisions in subsequent spiral structure extraction steps, and makes output information (such as spiral arm sizes and lengths) readily comparable across images.

Assuming that galaxy disks are roughly circular (a known property of galaxy disks that have not been perturbed by another galaxy [22]), the edge of the galaxy disk will appear elliptical when viewed at an angle, and will appear circular if and only if viewing angle effects have been removed. Thus, once the elliptical outline of the disk is known, it is possible to remove image variation due to viewing angle, centering, and apparent size by stretching the image so that the ellipse becomes a circle, cropping the image to a square with the circle inscribed, and resizing the image to a standard resolution (256×256 in our case, but other resolutions would likely work as well). Several methods are available to determine the elliptical outline of a galaxy (explicitly or implicitly) and to de-project galaxy images (e.g., [23], [21]). Here, we describe a method that is tailored to our purposes (it avoids calculating unnecessary information, is widely applicable across images, does not need manual guidance, and retains sufficient accuracy), but any method that can detect the elliptical outline in an automated manner (i.e., with sufficient speed and reliability) can be used.

We detect the elliptical outline by iteratively fitting a two-dimensional Gaussian because this can be done without any manual supervision, and does not waste computation time in determining unnecessary information (we only need the elliptical outline and not, for example, the exact brightness concentration of galaxy light in the disk and bulge). It also allows the use of images already processed for human viewing (e.g., JPEG, PNG) in addition to FITS images. For FITS images, we apply the brightness transformation described in Section 3.2. Otherwise, we skip brightness transformation and use the image as is. The two dimensional Gaussian function is not a perfect description of the light distribution of galaxy disks, but it is close enough to get the elliptical outline of the disk, which is all we need. Additionally, given weights for each pixel (e.g., as calculated from a previous iteration), the optimal values for the Gaussian parameters (the mean and covariance) have a closed-form solution, avoiding an expensive optimization at each iteration. Using an iterative fit confers resistance to non-galaxy light sources such as foreground stars and imaging noise; these objects can be successfully avoided in many cases (see Figure 3.4 for an example). There are a few remaining cases where this resistance is not enough, but see Section 3.4 for an optional addition that can complement the resistance provided by iterative fitting.

At each iteration of the Gaussian fit, we compute the weighted mean and weighted covariance of the Gaussian. The pixel weights for iteration k are calculated as

$$w_{ij}^{(k)} = \begin{cases} I_{ij} \cdot \frac{1}{|I|} & k = 0 \\ I_{ij} \cdot \mathcal{N}(\begin{bmatrix} i \\ j \end{bmatrix} | \mu^{(k-1)}, \Sigma^{(k-1)}) & \text{otherwise} \end{cases} \tag{3.3}$$

Where I_{ij} is the image intensity (after brightness transformation) at position $\begin{bmatrix} i \\ j \end{bmatrix}$, $|I|$ is the number of pixels in the image, $\mu^{(k-1)}$ and $\Sigma^{(k-1)}$ are the weighted mean and weighted covariance from iteration $k-1$, and $\mathcal{N}$ is the standard Gaussian density function in two dimensions, i.e.,

$$\mathcal{N}(\begin{bmatrix} i \\ j \end{bmatrix} | \mu^{(k-1)}, \Sigma^{(k-1)}) = \frac{\exp\left\{-\frac{1}{2}(\begin{bmatrix} i \\ j \end{bmatrix} - \mu^{(k-1)})^T \frac{1}{\Sigma^{(k-1)}}(\begin{bmatrix} i \\ j \end{bmatrix} - \mu^{(k-1)})\right\}}{(2\pi)|\Sigma^{(k-1)}|^{\frac{1}{2}}} \tag{3.4}$$

With these weights, we use the standard formulas to calculate the weighted mean and weighted covariance, i.e., at iteration k the weighted mean is calculated as

$$\mu^{(k)} = \frac{\sum\limits_{(i,j)} w_{ij}^{(k)} \cdot \begin{bmatrix} i \\ j \end{bmatrix}}{\sum\limits_{(i,j)} w_{ij}^{(k)}} \tag{3.5}$$

and the weighted covariance is calculated as

$$\Sigma^{(k)} = \frac{\sum\limits_{(i,j)} w_{ij}^{(k)}(\begin{bmatrix} i \\ j \end{bmatrix} - \mu^{(k)})(\begin{bmatrix} i \\ j \end{bmatrix} - \mu^{(k)})^T}{\sum\limits_{(i,j)} w_{ij}^{(k)}} \tag{3.6}$$

In practice, we add a small value ϵ to the diagonal of $\Sigma^{(k)}$ in order to avoid problems with numerical stability. We set $\epsilon = 2^{-52}$, i.e., the distance from 1.0 to the nearest double precision floating point number. In a large majority of cases, adding ϵ to the diagonal is not needed, but since the other effects of ϵ are negligible, we always use it.

For several iterations, the Gaussian contours will become a progressively better fit to the galaxy disk. After that, the fit will temporarily plateau at the disk, and then the Gaussian contours will shrink further to fit the galaxy bulge. Since we are looking for the outline of the galaxy disk (not the bulge), we must detect the temporary plateau at the galaxy disk and stop the iterations at this point. We note that $w^{(k)} = \sum\limits_{ij} w_{ij}^{(k)}$ approximately measures how well the Gaussian parameters match image brightness values because $w_{ij}^{(k)}$ is a product of image intensities and Gaussian density values, with the image intensities fixed and Gaussian density values normalized so that they sum to 1. Thus, high values of $w^{(k)}$ indicate that the Gaussian density values are distributed to match bright parts of the image. We can then detect the plateau as the point where $w^{(k)}$ is improving but at a lower rate (indicating the plateau), with this slowdown in improvement small and getting smaller (indicating that the fit is about to exit the plateau). We can quantify these criteria with the second derivative of $w^{(k)}$; specifically, we define

$$c^{(k)} = w^{(k)} - 2 \cdot w^{(k-1)} + w^{(k-2)}, \quad k \geq 3 \tag{3.7}$$

and stop the Gaussian fit at iteration k if $k > 3$, $0 > c^{(k)} \geq -\delta$, and $c^{(k)} \geq c^{(k-1)}$, using an empirically determined value of $\delta = 0.005$. As extra protection against pathological cases, we can also set a maximum number of fits, stopping (and recording a warning) if $k \geq 25$, but in practice this is not needed. After the Gaussian fit is stopped at the plateau, the 10^{-9} contour of the Gaussian likelihood consistently traces the outline of the galaxy disk (except possibly when bright foreground stars are present, but these can usually be ignored with the iterative fitting, and otherwise counteracted using the methods described in Section 3.4). Figure 3.4 illustrates this iterative fitting, and its termination at the plateau. After the elliptical outline of the disk has been determined, the image can be transformed so that the galaxy appears circular, centered, and at a constant resolution, as mentioned earlier. Figure 3.5 provides several examples of this standardization.

Two variations are possible in the determination of the galaxy center during the ellipse-outline fit. In some circumstances, the galaxy center will be known in advance (such as if the galaxy has already been centered in the image, as it is in many data sources), in which case we can treat μ as fixed, and only calculate $\Sigma^{(k)}$ and $w_{ij}^{(k)}$ during each iteration. When it is necessary to determine the center, we can improve its precision by noting that the Gaussian fit will be better centered over the galaxy center once the Gaussian passes

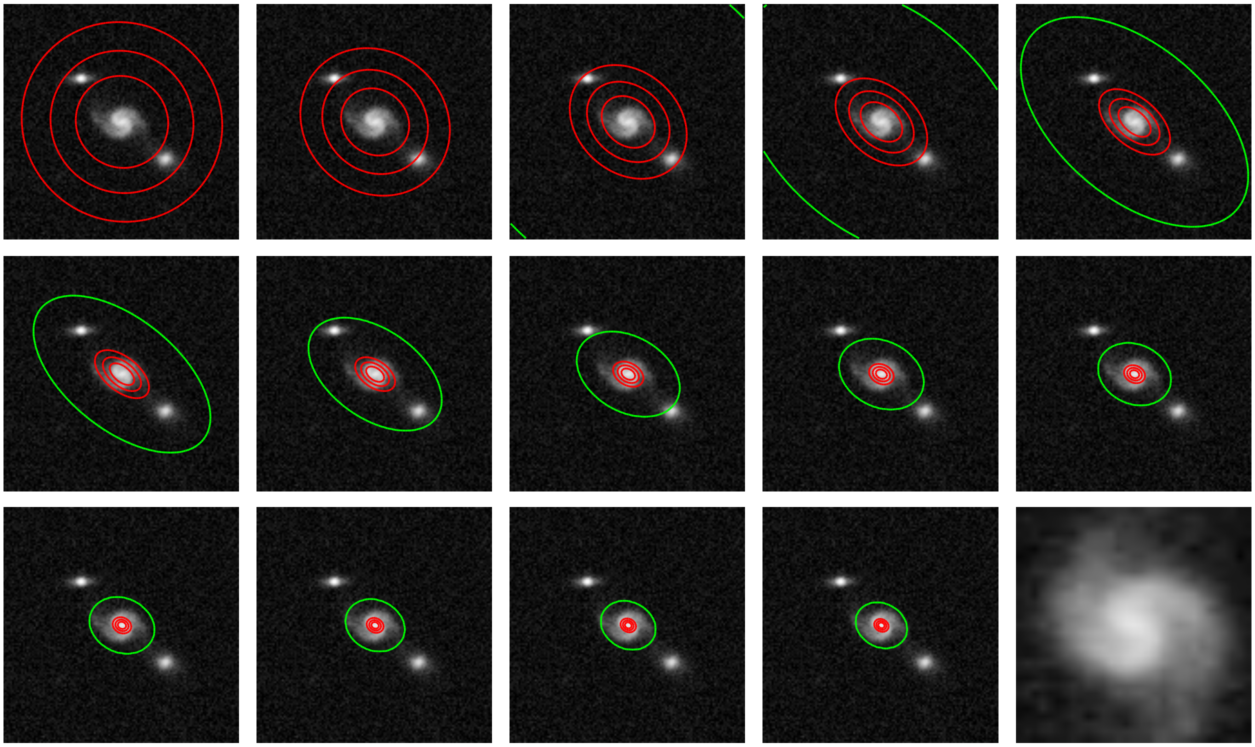


Figure 3.4: Images are standardized to a consistent viewing angle and size by detecting the elliptical outline of the galaxy disk and then transforming the image so that the ellipse appears as a circle inscribed in the image (which is resized to a constant resolution). The elliptical outline is determined by iteratively fitting a two-dimensional Gaussian and then stopping when the fit plateaus (to prevent the fit from collapsing around the galaxy bulge). The 14 iterations needed for this example image are shown from left to right and then top to bottom. For each iteration, three characteristic Gaussian likelihood contours are shown in red, along with the contour used for the galaxy disk outline, in green. Since the other objects in the image all have lesser total image intensity than the target galaxy, these other objects are successfully ignored. The final standardized image is shown at the bottom right.

the plateau and shrinks until it is only fitting the galaxy bulge. We can take advantage of this by performing the iterative Gaussian fit twice. The first time, we stop when either $\sqrt{(\mu^{(k)} - \mu^{(k-1)})^T(\mu^{(k)} - \mu^{(k-1)})} < \tau$ (indicating convergence of the center because the distance between successive μ values is small) or $k \geq 100$. The convergence threshold τ has an empirically determined value of 0.02. After this fit has finished, we record the final value of μ. In the second Gaussian fit, we fix μ to the value found previously.

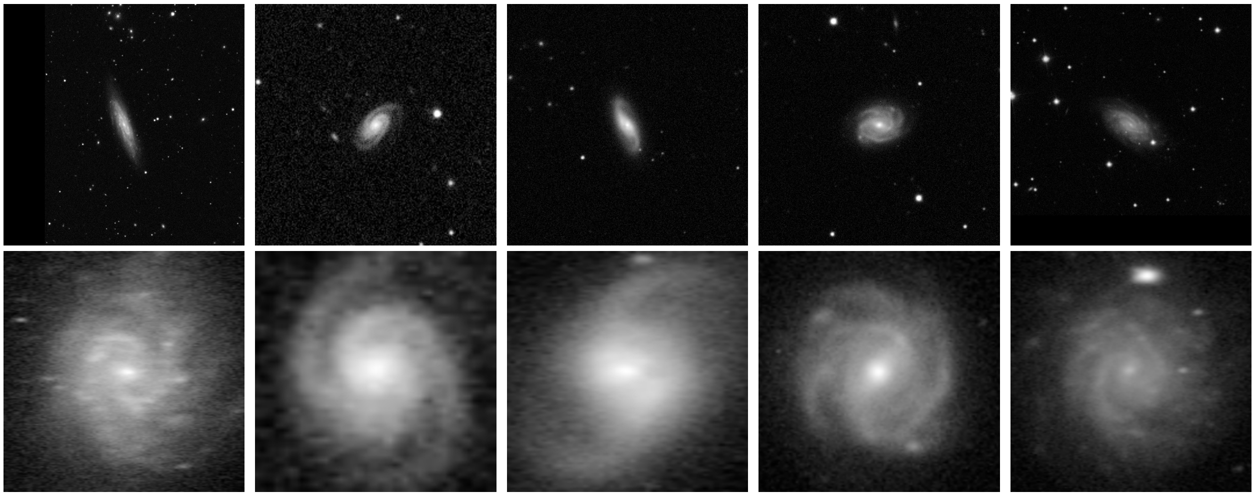


Figure 3.5: Variations due to galaxy viewing position, particularly in viewing angle and apparent size, are removed using the ellipse fit to the galaxy disk (see text and Figure 3.4). The top row displays images before this standardization, and the bottom row displays the corresponding images after standardization.

3.4 Foreground Star Removal

When determining the elliptical outline of the galaxy disk for image standardization, as discussed in Section 3.3, the iterative fitting often but not always resists disruption from stars (or other objects) in the image of target galaxy. When the ellipical-disk-outline fitting does fail, it tends to do so by either finding the outline of a star rather than a galaxy, or by straddling the galaxy and another object. Often, star disruption is not a problem, because astronomers have methods for detecting and masking foreground stars. However, star masking is often done separately from the sky surveys that acquire galaxy images, and to the best of my knowledge there does not appear to be a consistently used, detailed specification of how stars are masked and removed. Therefore, I have developed a star masking process that can be used in situations where star masking is not otherwise available. This star masking requires images to be available in FITS format, but this format, or something convertible to it without loss of information, is almost always used for astronomical images. PNG or JPEG images can still be used if their underlying FITS images are available, in which case the FITS image is used to obtain the star mask, and then the PNG or JPEG image is used instead of the brightness transformation from Section 3.2.

My approach follows the general idea of other star masking procedures in that it uses and cleans the output of SExtractor (an abbreviation of Source Extractor) [7]. SExtractor detects pixel regions corresponding to different light sources, and attempts to classify these regions as stars versus galaxies (i.e., point sources versus non-point sources). However, these are challenging tasks, in part because of image noise and other imperfections in the image, because galaxies are composed of many light sources, because many pixels convey brightness from more than one light source, and because of variation in the imaging environment, such as the filter bandpass and atmospheric conditions. As such, we cannot expect SExtractor output to be perfect for our purposes. In particular, we cannot be sure whether a detected region is a star or a galaxy (without per-dataset human judgment on SExtractor parameters and classifier training, which can be done partially in place of the methods of this section, but which we want to avoid here for generality and automation), there may not be a one-to-one correspondence between SExtractor-detected regions and objects of interest in the image (in particular, the galaxy may be broken into multiple regions, especially since galaxies often contain multiple resolvable light sources), and the shapes of the detected regions may be highly ragged, non-contiguous, and otherwise irregular. To address these issues, we clean the SExtractor output to make it more amenable to star masking (though not necessarily more suitable for other applications of SExtractor that do not concern us here), and we only use star masking when and to the extent necessary. To achieve the latter, we apply star masks of increasing aggressiveness (initially trying not to use a star mask at all), and only use star masking in steps where it is needed. The primary use of this star masking is to aid the galaxy disk outline determination in Section 3.3, but star masking can also be used to reduce the impact of foreground stars during the image brightness transformation step (Section 3.2).

The first step in using SExtractor is to run it under its default configuration (using the configuration files provided with the executable), except that we ask for a segmentation image as output (by specifying `CHECKIMAGE_TYPE SEGMENTATION` on the command line or in the parameter file). We used SExtractor version 2.8.6, but other versions may work as well. The SExtractor segmentation image gives the identifier of the SExtractor region assigned to each pixel. We then refine this segmentation image before using it to generate a set of star masks. First, we remove holes in the pixel regions, which restores individual pixels and small groups of pixels omitted due to noise, and subsumes objects contained entirely within other objects. The latter is less common; when it does occur, it can combine regions corresponding to the same object (which is often a galaxy), but can also lead to stars being included in the galaxy region. Including stars in the galaxy region has not been a significant problem; if the star is bright enough to disrupt the ellipse fit to the galaxy, the star region will generally be large enough that it will not be fully contained within the galaxy region. After this hole-filling, we replace each pixel's region value with the majority vote within the 3x3 neighborhood centered on the pixel. Ties can occur at image borders, and are settled in favor of the current region value. For each region, we then keep only the largest connected component (using a 3x3 neighborhood), which removes small, often-noisy patches (which are reassigned to the background).

After cleaning the SExtractor output, we use it to produce star masks with increasing levels of aggressiveness. The first "mask" is trivial (it does not mask anything). For the non-trivial star masks, we must first determine which SExtractor region corresponds to the galaxy (or the most prominent part of it, if SExtractor broke the galaxy into multiple regions). We call this region the "primary region." Given the difficulties SExtractor faces in producing perfectly reliable output (for our purposes) without dataset-specific tuning, combined with the fact that astronomers can (almost always) automatically center the galaxy of interest (or at least place it closer to the center than any other object), we simply take the region assigned to the image-center pixel as the primary region. If no region is assigned to the center, we instead use the region with a pixel closest to the center. Next, since regions contiguous to the primary region may be part of the galaxy (which would be, for our purposes, a false separation by SExtractor) or may be another object such as a star, we also define an "extended region" consisting of the primary region and all regions directly or transitively adjacent to the primary region. The raggedness of SExtractor regions can interfere with region adjacency, but this was counteracted by the region-cleaning steps mentioned above. We then have four types of pixels for defining star masks: background pixels, pixels in a SExtractor region that is not part of the primary or extended region, pixels that are part of the extended region but not the primary region, and pixels that are part of the primary region.

We can then use this information to mask out stars that interfere with the elliptical-disk-outline determination described in Section 3.3. To do so, while only using star masking where and to the extent necessary, we must determine when stars are disrupting this disk-outline determination process. Although star disruption is rare, it is severe when it occurs (if a star caused the ellipse fit to only be partially distorted, this distortion can be corrected in subsequent iterations). Thus, star disruption tends to be easily detectable as long as there is at least a weak sense of what the ellipse fit would be like without star disruption. Since astronomers can at least approximately determine galaxy centers automatically, we can detect star disruption by looking for ellipse centers that drift too far from the image center.[3] In particular, we consider stars to be disrupting the ellipse fit when, after performing this fit, the Euclidean distance between the ellipse center and the image center is greater than 2.5 pixels. The exact choice of this threshold is not critical. Since star disruption tends to be severe (when it occurs at all), its effect on ellipse centering is severe as well (the center tends

---

[3] In an earlier attempt to detect disruption from stars, the ellipse was divided into inner and outer parts, and then the ratio of the two average brightness values was calculated. Since stars often cause the ellipse fit to straddle the star and the galaxy, the outer part of the ellipse would be unusually bright, and the inner part would be unusually dim (since there would not be an object in the middle). However, this often resulted in false detections for galaxies with dim bulges and/or a foreground star near the galaxy edge (even a small star that would not disrupt the ellipse fit), and did not detect star disruption when the ellipse fit focused entirely on a star. Thus, we found the ellipse center distance to be a better means of detection, to the point where it was worth introducing the assumption that the galaxy is close to the image center, especially since this assumption is reliable in practice.

to move to a foreground star, or between a galaxy and a foreground star), so star-disruption cases can usually be easily distinguished from normal cases. To handle the rare instances of ambiguous distance values, we set a tight threshold because the risk of using an unnecessarily aggressive star mask is preferable to the risk of choosing the wrong elliptical image region, especially since an ellipse fit around the primary region is usually reasonable and the star mask can be removed after the ellipse fit.

When the ellipse fit is considered to be disrupted by a star, we use the next (more aggressive) star mask and repeat the ellipse fit. We first try the ellipse fit without a star mask. If star disruption is detected, we then try the ellipse fit with a star mask that zeros pixels in all SExtractor regions that are not part of the primary or extended region. This removes stars and other objects that are very unlikely to be part of the galaxy (due to their distance in the image), but are sometimes bright enough to cause a problem. If star disruption is still detected, the next star mask zeros pixels in all SExtractor regions except the primary region and the background (in particular, this zeros out the extended region). This can remove stars that are closer to the galaxy, at the risk of masking out part of the galaxy (though even if this happens, the ellipse fit often succeeds anyway, and subsequent steps see the full galaxy because the mask is removed after the ellipse fit finishes). If star disruption is still detected, we zero out all pixels outside the primary region, regardless of whether these pixels are background or part of a SExtractor region. This is rarely needed, but it can cope with missed SExtractor detections in the unlikely case that they exist. This could also cope with bright pixels that were assigned to the background during the SExtractor region cleaning (e.g., removal of non-contiguous pixels), but this case has never been observed, perhaps because SExtractor would likely consider an image-separated bright pixel region as a separate object detection. If all of these masks fail to adequately reduce the distance between the ellipse center and the image center, the ellipse fit from the final star mask is used. The level of star masking needed can be reported to the user (in case it is important to verify the rare cases where aggressive star masks were needed, for example), and the image standardization process is now more resistant to disruption from stars. Figure 3.6 provides examples of the different levels of star masking that can be applied.

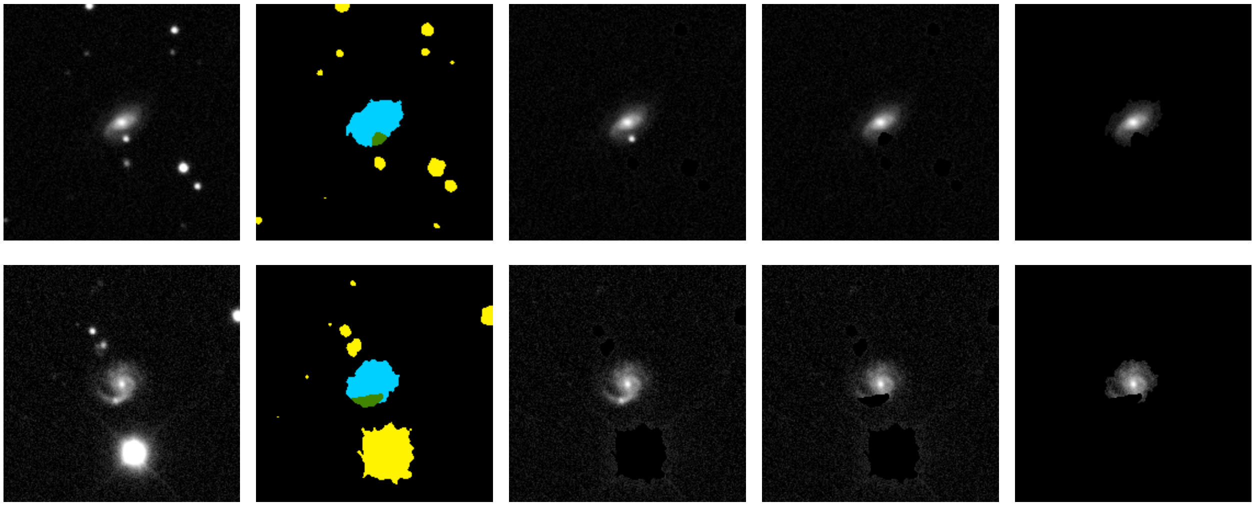


Figure 3.6: Star masking can prevent stars (and other objects in the image) from interfering with the image standardization described in Section 3.3, and can improve the consistency of the brightness transformation described in Section 3.2. The first column gives example images that contain objects other than the galaxy of interest; the image intensities are transformed (Section 3.2) for visibility. The second column shows image regions produced using cleaned SExtractor output (see text). 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. The second level also removes pixels that are in the extended region but not the primary region. 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.

If an input image has a severe amount of CCD noise, some pixels in the “empty” regions of the image may have unusually bright values after the brightness transformation described in Section 3.2. If this effect is strong enough, brightness values comparable to the galaxy disk can be scattered throughout the “empty” image regions, causing the ellipse fit to stop while it still encompasses an area much larger than the galaxy disk. We can address this situation with a procedure similar to star masking. If imaging noise is affecting the ellipse fit, the contours of the ellipse will typically extend beyond the image region. When this occurs, we can (during ellipse fitting only) zero out pixels that are not part of the primary or extended region, and then repeat the ellipse fit. This is usually sufficient to remove imaging noise, but if the ellipse fit still extends outside the image boundary, we can again repeat the ellipse fit, this time also zeroing out pixels in the extended region (but not the primary region). If the ellipse still extends outside the image boundary, the last ellipse fit is used. The level of "noise masking" used is also reported to the user; if all noise masking measures fail, it is likely that the field of view in the input image did not encompass the entire galaxy. This noise masking is rarely needed for the images we used, but it nevertheless increases the robustness of the ellipse-fitting procedure.

Star masking can also improve the consistency of the brightness transformation described in Section 3.2. In this brightness transformation, the β parameter controlled the transition point between near-linear and near-logarithmic treatment of the brightness values, and was chosen based on the distribution of brightness values (by computing a percentile). Foreground stars can distort this distribution by introducing additional extreme brightness values that would otherwise be rare. Since we are only interested in brightness values within the galaxy, we can avoid star-induced distortions to the brightness distribution by measuring the brightness percentile within the galaxy region rather than within the image as a whole. For this purpose we find it sufficient to use the primary region: even if SExtractor fragments the galaxy into more than one region, the primary region covers most of the galaxy, including the bulge and most of the disk. The lower part of Figure 3.6 gives an example of how the primary region covers most of the galaxy (in the upper example, the galaxy is not fragmented). This coverage is sufficient to get a reasonable brightness percentile value. Furthermore, losing some "typical" galaxy pixels from the percentile calculation is better then introducing spurious extreme brightness values (a risk incurred if other, possibly star-containing, SExtractor regions are included in the percentile calculation). In addition to modifying the selection of the β parameter, we clip image brightness values that exceed the maximum brightness found within the primary region (the region that contains the galaxy bulge, and thus the brightest part of the galaxy). This avoids wasting part of the image brightness range on values only seen from foreground stars. Even with these modifications, we still need the brightness transformation because the galaxy bulge also has extreme brightness values (so the spiral arms often occupy a small part of the brightness range even without foreground stars), but the star masking eliminates two sources of unwanted variation in the brightness transformation: non-galaxy influences on the brightness distribution used to select β, and the portion of the image brightness range occupied by the galaxy.

3.5 Orientation Field Generation

Given an image with transformed brightness (Section 3.2) and standardized apparent size and viewing angle (Section 3.3), we would like to determine pixel regions corresponding to spiral arm segments. However, the brightness values themselves do not give a strong indication of whether or how the pixel is part of a spiral arm. Spiral arms are generally neither the brightest nor the dimmest region of the galaxy and can vary in relative brightness compared to the rest of the galaxy components. Instead, the pattern of brightness is often much more informative. Thus, we calculate pixel-level features that capture some of these patterns, thereby providing better information about spiral arm membership. We do so by calculating an *orientation field* as described in [2, 3]. We summarize this procedure, and then describe a method for enhancing the image so that spiral arms are more distinctive in the orientation field.

3.5.1 Determining the Orientation Field Vectors

This orientation field generation procedure [2] uses the fact that, if a pixel is part of a spiral arm, then nearby pixels along the spiral arm (and the pixel itself) will generally be brighter than nearby pixels not along the spiral arm. Within a small region around the pixel, this will manifest itself as a roughly linear brightness pattern along the spiral arm. For each image pixel, the orientation field determines the strength and direction of this linear brightness pattern. Generating the orientation field has three main steps: applying oriented image filters at a given resolution to get per-pixel orientation strengths and directions, combining this orientation information across three image resolutions, and applying normalization and de-noising to the resulting orientation field. We summarize these steps here; for derivations and more detail, see [2].

For a single image resolution (scale), the initial orientation strengths and directions are determined by applying a set of oriented filters to an image via convolution. The convolution operation replaces each pixel with the weighted sum of its neighbors (including the pixel itself), with the set of weights given as the image filter.[4] The image filters are designed to favor linear patterns with (or close to) a particular angle θ. Several such filters will be constructed, one for each of a set of values of θ. The filters are based on the one-dimensional equivalent of the Laplacian-of-Gaussian function, i.e.,

$$h(x) = \frac{2}{\sqrt{3}}\pi^{-1/4}(1 - x^2)e^{-x^2/2}. \tag{3.8}$$

This function is illustrated in Figure 3.7a. When extended into two dimensions such that x is the distance from a line with angle θ, this produces a ridge function that, when used to produce pixel-neighborhood weights, will favor linear structure in the direction of θ. Since we only want to look for structure within a small neighborhood around the pixel, the weight

---

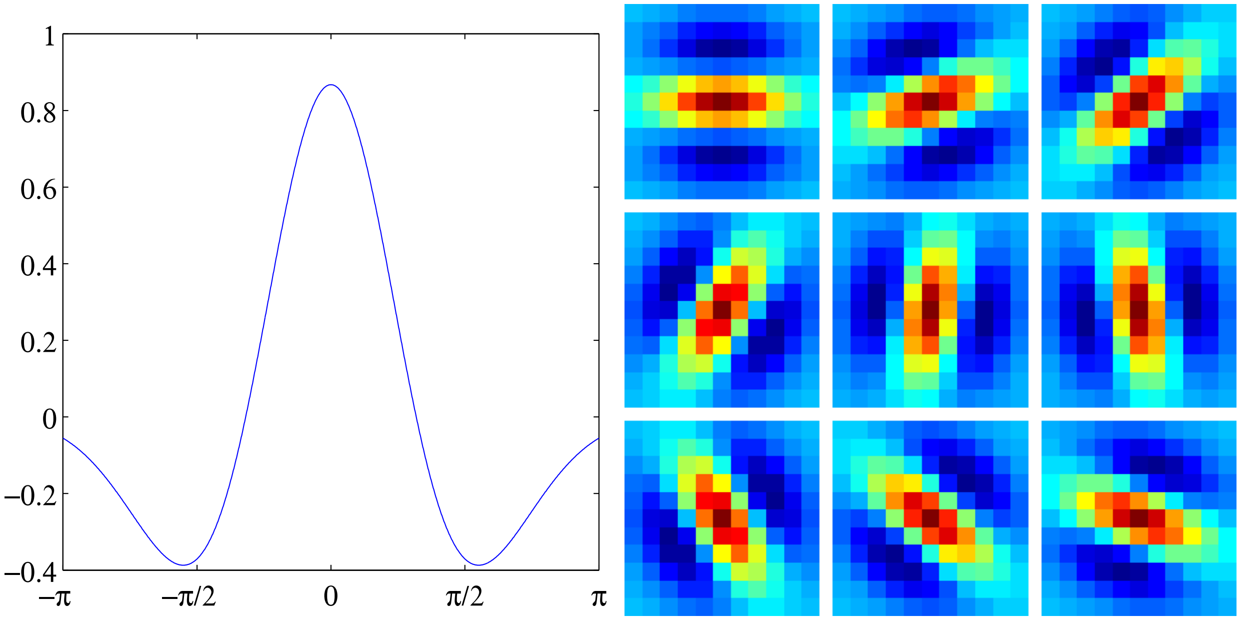
[4] The filter is first flipped in both dimensions, but this does not change the filters used here.

strengths are reduced according to their distance from the center of the image filter, with the fall-off determined using a Gaussian function. In [2], the reported filter function $\psi_\theta(x, y)$ applies this Gaussian attenuation isotropically. However, in consideration that $h(x)$ already falls off to near zero at the endpoints of the x range used in the filter, and to match the filter matrices given in Appendix F.2 of [2], we only apply the Gaussian attenuation along the line with direction θ. Each orientation filter is then produced using the function

$$\psi_\theta(x, y) = h(p_{\theta+90^\circ}(x, y)) \times \frac{1}{\sqrt{2}} e^{-\frac{1}{2}(p_\theta(x,y)^2)}, \tag{3.9}$$

where $p_\theta(x, y) = x \cos\theta + y \sin\theta$, and $h(x)$ was defined in Equation 3.8. The orientation filters are 11×11 pixels in size, with values produced from $\psi(x, y)$ such that the x and y values both have the range $[-\pi, \pi]$, with these endpoints corresponding to the outer edges of the filter pixels. The filters are then normalized (using the Euclidean norm). Nine filters are used, using angles $\theta = 0, \frac{\pi}{9}, \ldots \frac{8\pi}{9}$. These filters are illustrated in Figure 3.7b; note how they favor linear structure along the corresponding angle θ.

After applying image convolution with these filters (weights), the per-pixel orientation information is summarized as $\vec{w} = \sum\limits_\theta E_\theta \, e^{\imath \cdot 2\theta}$, where $\imath = \sqrt{-1}$ and E_θ is the square of the convolution result (using the filter corresponding to angle θ) for the pixel in question. The orientation angle and strength at each pixel are then given via the complex argument and modulus as $\arg(\vec{w})/2$ and $\|\vec{w}\|$, respectively. The orientation angle gives the direction at which the orientation filter response is the strongest. Angles differing by a multiple of π radians are equivalent. The orientation strength is the extent to which the filter responses are larger near this dominant orientation. This measure gives strong values to spiral arm regions (because these regions have a linear brightness pattern and thus have a strongly dominant orientation), while giving weak values to symmetric light sources (such as stars) and constant-brightness regions.



(a) The Mexican hat function (b) Orientation filters

Figure 3.7: Orientation-sensitive filter generation as adapted from [2]. The filters use the Mexican hat (one-dimensional equivalent of the Laplacian of Gaussian) function, left. This function is extended along one of nine directions, and then a Gaussian window is applied, producing the nine filters shown at the right. The strongest positive values are in red and the strongest negative values are in blue. Each filter yields increased responses when larger brightness values are aligned or nearly aligned with the filter's orientation, and yields reduced responses when larger brightness values are found in other directions (or when strong brightness values are not present at all). Combining the results of these nine filters produces a per-pixel measure of orientation strength and direction. Determining this orientation information at each pixel produces an orientation field.

At a single image resolution, this scheme assumes that the widths of the spiral arms roughly match the width of the ridge used in the filter. Along with the image standardization described in Section 3.3, which reduces variation in arm width, the need to match filter-ridge and arm widths can be at least partially addressed by combining orientation information from multiple image resolutions. Following the method described in [2], orientation information from two image resolutions can be combined as

$$V_{i,j}^{(f)} \leftarrow V_{i,j}^{(c)} + \frac{S_{i,j}^{(f)}}{S_{i,j}^{(c)} + S_{i,j}^{(f)}} (V_{i,j}^{(f)} - V_{i,j}^{(c)}). \tag{3.10}$$

Here, $V^{(c)}$ and $V^{(f)}$ are matrices expressing the orientation information from the coarser and finer scale, respectively, such that $V^{(c)}$ has been upscaled to the same dimensions as $V^{(f)}$ using interpolation operations (as one would do with image resizing), and $V^{(\cdot)}_{i,j}$ is a two dimensional Cartesian-coordinate vector giving the orientation at position (i, j). $S^{(c)}_{i,j}$ and $S^{(f)}_{i,j}$ give the orientation strength for position (i, j) using orientation information at the coarser and finer scale, respectively. Like $V^{(c)}$, $S^{(c)}$ has been upscaled to the same size as $S^{(f)}$. To combine orientation fields from more than two resolutions, we first perform this procedure for the orientation fields with the lowest and second-lowest resolutions. While there are orientation fields at higher resolutions, one can then treat the previous result as the new coarse orientation field, and repeat Equation 3.10 with the orientation field at the next-highest resolution. As in [2], we combine orientation information from three resolutions: the highest resolution (in our case, this is the standardized image resolution as discussed in Section 3.3), along with half and a quarter of this resolution.

After combining the orientation information from multiple image scales (resolutions), [2] performs a normalization and de-noising procedure. For each orientation field position, this procedure examines neighbors that are five pixels away in each direction along each axis, i.e., neighbors with positions

$$R(i, j) = \{(i-5,\ j), (i+5,\ j), (i,\ j-5), (i,\ j+5)\}. \tag{3.11}$$

The orientation strengths are then updated as

$$S_{i,j} \leftarrow \operatorname{median}\left(\left\{\frac{\max(|V_{i,j} \cdot V_{i',j'}| - \cos(\pi/4), 0)}{\left\|V_{i,j}\right\| \left\|V_{i',j'}\right\|} \ :\ (i', j') \in R(i, j)\right\}\right). \tag{3.12}$$

This expression performs normalization to make the orientation strengths more invariant to the overall brightness of the region (otherwise, uniformly increasing the brightness of the region would increase the orientation strength, and orientations would often be stronger near the galaxy bulge). It also zeros out noisy orientation vectors, under the assumption that regions without an orientation pattern produce orientation vectors with low strength and random direction (particularly random direction that varies among nearby pixels). The neighborhood distance of five pixels was chosen because it is close enough for neighborhood-related expectations to hold, and far enough for orientation vector alignment to occur for reasons other than the smoothing effect of the orientation filters.

3.5.2 Enhancing Sensitivity to Spiral Arms via the Unsharp Mask

When producing the orientation field as-is in practice, the orientation field directions would "flow" along spiral arms, but the orientation strengths would usually be similar for arm regions and non-arm regions, even for well-resolved galaxies.[5] When using the original FITS images (with or without subtracting the minimum), this filled non-arm regions with spurious orientation vectors of high strength. When rescaling the images to a $[0, 1]$ range (with or without the brightness transformation given in Section 3.2), very few orientation field vectors (if any at all) remained after normalization and de-noising (Equation 3.12). Comparing the orientation fields before and after the adjustment from Equation 3.12, it was found that either the normalization or the de-noising tends to overpower the other. The normalization can be helpful (and sometimes necessary) for bringing out spiral arms, but when it does so, it also greatly amplifies what would otherwise be much smaller orientation strengths in the non-arm regions. Alternately, when the de-noising zeros out noisy orientation vectors, it also tends to zero out orientation vectors along spiral arms.

The presence of many strong orientation vectors in non-arm regions may not be problematic for the symmetric two-arm spiral model in [2], due to the restrictions that this model imposes. For example, if the orientation field has three regions with strong, locally consistent

---

[5] We also tried applying the normalization before subtracting the cosine in Equation 3.12 (since the cosine subtraction would then be applied to a normalized dot product) to ensure that this was not the intended equation in [2], but this change did not improve the situation.

orientations, the region conforming the least to the symmetric two-armed spiral model can simply be treated as noise or otherwise ignored. However, our list-of-arcs model (Section 3.1) is far more general than the symmetric two-arm spiral model, and if the third region is also a legitimate spiral arm region, we want to report it as such. Consequently, our method cannot be robust to spuriously strong orientations in the same way as [2]. Instead, we must either find a way to determine which orientation field vectors truly come from spiral arms, or suppress spurious orientation field vectors during the orientation field calculation process. The former would introduce additional uncertainty and require additional computation time. Furthermore, we aim to apply our method to a large set of images, and many of these images are less well resolved than the ones generally used in [2], so the difference between arm orientations and non-arm orientations would become even less well defined. Fortunately, we can instead make the orientation field generation procedure more sensitive to arm regions and less sensitive to non-arm regions. This is accomplished by using an unsharp mask operation [35] (a contrast enhancement, despite its name) to the standardized image before orientation field generation.

When applying the unsharp mask, the image is updated as

$$I \leftarrow I + \alpha(I - G * I) \tag{3.13}$$

where $*$ is the convolution operation and G is a matrix of Gaussian values (i.e., the filter used in convolution). This amplifies differences between the image and a blurred version of this image. Consequently, pixel values are increased to the extent that their neighbors are dimmer, and decreased to the extent that their neighbors are brighter. The values of I, which were originally in the range $[0, 1]$, are then clipped back to this range.

If the scale (i.e., the standard deviation) of the Gaussian blur is large enough to include mostly non-arm pixels in the neighborhoods of arm pixels, then the brightness of spiral arm pixels will be increased because arms are brighter than most or all of their surrounding non-arm regions (most surrounding pixels are from the galaxy disk; a few may be from the bulge). Likewise, in the inter-arm regions (i.e., where spurious orientation vectors were problematic), the brightness will be reduced, especially near the spiral arms. This causes the light from the spiral arms to better match the pattern sought by the orientation filters. When the orientation of one of the filters (visualized in Figure 3.7b) aligns with the spiral arm at a particular position (and when the width of the filter is close to the width of the arm, which was already a requirement), then the filter weights with the highest positive values are matched to the spiral arm pixels, which now have increased brightness, and the filter values with the strongest negative values will be matched to pixels outside the spiral arms, which now have reduced brightness. Figure 3.8 illustrates this effect. Since the orientation filter response is enhanced to the extent that filters align with arm regions, it becomes easier for the orientation field de-noising procedure to zero out only non-arm regions.

As shown in Figure 3.9, sufficiently large scales of Gaussian blurring also cause the unsharp mask to function as a crude form of galaxy disk subtraction, since light from the disk has a lower spatial frequency than light from the arms, and using the large Gaussian scale subtracts light with the lower spatial frequencies. Fortunately, the arm-amplification and disk-subtraction interpretations of the unsharp mask agree that the Gaussian scale should be reasonably large, and due to the image standardization discussed in Section 3.3, we can use a constant scale. In particular, we found that using a Gaussian filter with standard deviation 25 satisfies the arm-amplification and disk-subtraction considerations. When performing the Gaussian blur, pixel values outside of the image should not be set to zero, since this would artificially reduce blurred values near the image border, and thus cause the unsharp mask to amplify pixels near the border even if they are not brighter than nearby regions within the image. Instead, pixel values outside of the image are determined by the value of the closest pixel within the image, which allows near-boundary pixels to be treated like the other pixels. Next, the unsharp mask strength α must be determined. Since the image brightness was

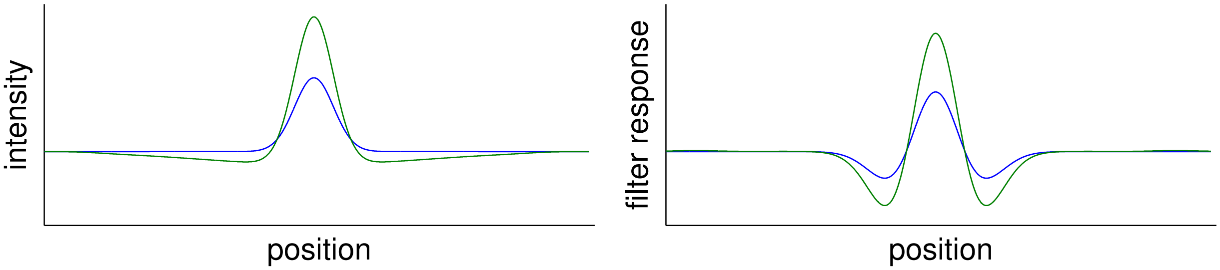


Figure 3.8: The unsharp mask amplifies the relationship between orientation filter responses and the extent to which the orientation filter is positioned over and aligned with a spiral arm region. The effects of filter position are shown here in one dimension, with a schematic view across a spiral arm. Values without the unsharp mask are displayed in blue, and values with the unsharp mask are displayed in green. Due to the contrast enhancement of the unsharp mask, image intensities (left) are increased within the relatively bright arm region and decreased nearby, making the arm more prominent. Comparing these changes to the orientation filter profile in Figure 3.7a, we note that when the orientation filter has approximately the same width as the spiral arm and the filter is centered within the spiral arm, image intensity values are increased where the filter values are the most strongly positive, and decreased where the filter values are the most strongly negative. Consequently, orientation filter responses (right) are increased when the filter center is within the arm region. Similarly, when the orientation filter is centered near but not within the arm region, its responses are reduced. In two dimensions, analogous affects are seen for orientation alignment.

transformed (see section 3.2) and can be rescaled to the constant range $[0, 1]$, we can use a constant value for α. In consideration of typical differences between arm and inter-arm brightness, as well as the tradeoff between arm brightness amplification and image clipping, we find a value of $\alpha = 6$ to be effective.

Although the unsharp mask amplifies arm brightness, it can also exaggerate noise, which can be especially problematic for relatively poorly-resolved galaxies. We can optionally reduce this noise by applying a median filter [28], which replaces a pixel's brightness value with the median of the neighborhood around a pixel. We use a 3×3 neighborhood, and apply the median filter directly before the image standardization in Section 3.3. By performing median filtering before resizing the image (in the image standardization step), the median filter size is consistent with respect to the input image resolution. If all of the input images have a consistent resolution in arcseconds per pixel (as they typically would if they were generated

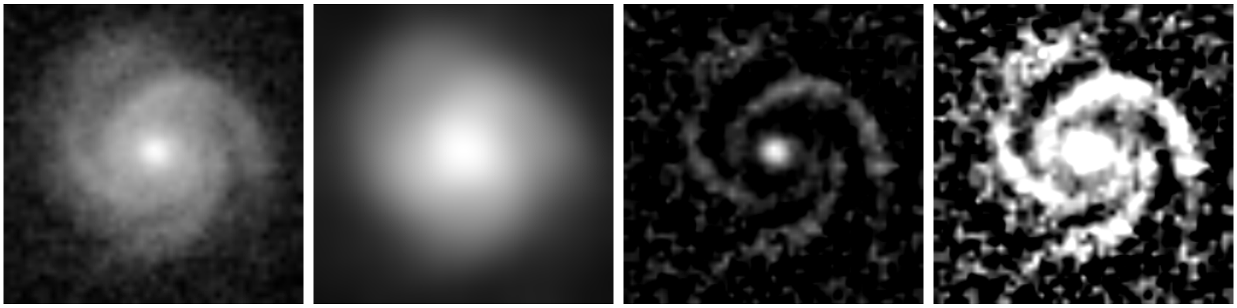

(a) (b) (c) (d)

Figure 3.9: The unsharp mask can function as a crude form of disk subtraction. Starting with an image of standardized viewing position (Section 3.3) as shown in (a), a sufficiently large scale of Gaussian blur will leave mostly disk features (b). When this blurred image is subtracted from the unblurred image, the arms become more prominent (c). The unsharp mask adds a multiple of this difference, amplifying the arm-prominence effect (d).

from a sky survey), then the filter size is consistent with respect to size on the sky, and if the images are provided at the original imaging resolution, then the median filter is applied at a point where much of the imaging noise affects pixels independently.

After unsharp masking, the orientation filters are more sensitive to spiral arm structure. High-magnitude orientation vectors still “flow” along the spiral arms, and the orientation vectors are now weak or non-existent in inter-arm regions. The resulting orientation field is shown in Figure 3.10. In subsequent steps, the orientation field will be the primary representation of the image, with one orientation field vector for each pixel. The image brightness values will not be needed further (except optionally as weights of least-squares fits to spiral arms).

3.6 Handling Bars in Spiral Galaxies

We are primarily interested in detecting arms in spiral galaxies, but some spiral galaxies also have bars (see Section 1.1 and Figure 1.2). Although bar information can be useful in astronomy, methods already exist for fitting bars (see Section 2.2). Thus, our primary

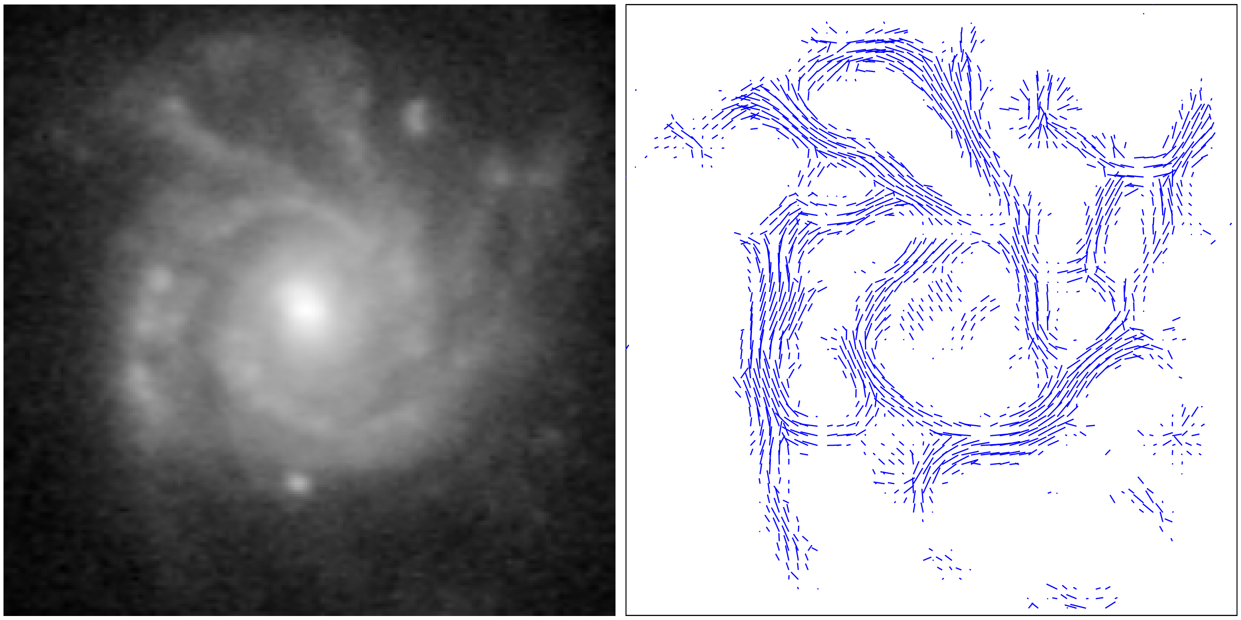


Figure 3.10: A standardized image (see Section 3.3), left, and its orientation field (after unsharp masking), right. For viewability purposes, only every fourth orientation vector (in each direction) is shown.

aim with bars is simply to make sure that they do not disrupt the spiral arm detection. Such disruption can occur because the orientation field generation procedure responds to bars in the same way as arms, since both are locally linear. Without accounting for bars, bar regions can then appear to be arm regions, introducing spurious arm segments into the spiral structure description. For many applications of our method this may not cause serious problems because the other arms will generally not be affected,[6] and bar-fitted arm segments have characteristics (e.g., low pitch angle, proximity to the image center, angular coverage near 2π radians, pixels in the image center) that could likely be detected in post-processing. Nevertheless, since bar detection (and removal) can still improve the accuracy of the spiral arm structure description, and since bar detection can benefit from orientation-field information that is not needed in the final output, we perform bar detection before looking for spiral arm-segment structure. This bar detection is not sensitive to all bars (for

---

[6] It is possible for arms adjacent to the bar to be affected because the bar could be misinterpreted as an extension of the arm, but such arms tend to visually join the bar at a sharp angle, so the arm and bar can be treated in the same way as arm "forks." As discussed in Section 3.7.2, these forks are split into separate regions.

example, since the bar is present near and on opposite sides of the galactic bulge, the bulge may overpower the bar to the point where the orientation field does not respond to the bar), but this is not a problem because we only need to detect bars that are prominent enough to have spiral-arm-like orientation patterns in the orientation field.

Taking advantage of the fact that the overall shape of the bar is linear, we apply the Hough transform [18], a procedure commonly used for line detection, in order to look for bars. The Hough transform parameterizes lines with a distance and an angle, since the slope-intercept form would be problematic for vertical line detections. The specified line is tangent to a circle, with the radius given by the distance parameter and the point of tangency given by the angle parameter (the Hough transform does not model or detect line endpoints). To find prominent lines, the Hough transform maintains an *accumulator matrix* of votes for all pairs of discretized values of the line parameters. Each pixel then votes for the parameter pairs of all compatible lines. The main challenge in our use of the Hough transform is that it will detect all collinear brightness patterns, not just the bar, so we must eliminate all of these non-bar line detections.

Many line detections can be eliminated because they are geometrically inconsistent with known properties of galactic bars. In particular, the linear pattern of the bar passes through the galaxy center, so we can eliminate any lines that do not come close to the galaxy center. Furthermore, we employ the assumption that bar length is nearly symmetric on both sides of the bulge. This comes at the expense of detections for hypothetical asymmetric bars, but the assumption is useful overall due to the overwhelming proportion of bars that are at least roughly symmetric. To apply the Hough transform in a way that favors these properties, we compute a set of accumulator matrices $A^{(r)}$. Each matrix $A^{(r)}$ includes only Hough transform votes within a circular region; this region has radius r and is centered on the galaxy bulge (as determined by the image standardization step in Section 3.3). These matrices can be computed as a single Hough transform (without changing the result) by choosing an appropriate pixel vote order. In the traditional Hough transform, pixels have a vote if they are considered an edge by an edge detection method. Here, the orientation field takes the place of the edge detection, and we use orientation strengths as pixel vote weights. In the Hough accumulator matrix, the distance parameter is binned in increments of 5 pixels, and the angle parameter is binned in increments of $\pi/360$ radians.

Each matrix $A^{(r)}$ is then given the score

$$\frac{\max\limits_{i,j}(A_{ij}^{(r)})}{r}, \tag{3.14}$$

with indexes i and j corresponding to the discretized distance and angle parameters in the Hough transform. We then select the radius with the highest score. The circular window for $A^{(r)}$ decides pixel voting eligibility in a way that favors lines that pass through the center (i.e., the galactic bulge), a line-placement property that we strongly expect from bars. This window also enables determination of the line (bar) endpoints, which the plain Hough transform does not detect. The $\max\limits_{i,j}(A_{ij}^{(r)})$ term favors strong line detections. The radius penalty normalizes against larger radii gathering more votes, and favors symmetric bars (asymmetric bars lose some votes on the shorter side while still incurring the full radius penalty). The largest eligible value for r is the image radius because this is the largest circular region that fits within the image. Since the galactic bulge often produces strong orientation field vectors radiating away from the center, the minimum value of r should exclude regions where the bulge produces strong linear votes in all directions with a small radius penalty. The presence of omnidirectional line votes can be quantified as $\max\limits_{i,j}(A_{ij}^{(r)})/\min\limits_{j}(\max\limits_{i}(A_{ij}^{(r)}))$ close to 1, since the indexes i and j correspond to values for the distance and angle parameters of a line. We then set the minimum value of r to the point where this ratio falls below an empirically determined threshold of 1.5.

The highest-scoring value of r gives the radius of the line segment corresponding to the strongest bar detection. The angle of this line is then the angle parameter of the Hough transform bin with the highest vote, i.e, the angle parameter with index $\underset{j}{\operatorname{argmax}}(\max_i(A_{ij}^{(r)}))$. Since we calculate this highest-scoring bar detection for all images, but not all galaxies have bars, we must determine whether the strongest bar candidate actually corresponds to a galaxy bar. Higher values of the score in Equation 3.14 indicate stronger votes for a linear pattern in the orientation field, so bar presence is more likely for high values of this score. If a bar actually exists, there should also be strong contrast between the brightness along the bar and the brightness across the bar. To test this, a second Hough transform is computed with the image intensities as vote weights, using a circular region with radius r (determined from the previous step), and producing accumulator matrix B. The image intensities come from the image standardization step (Section 3.3), except that the image is not stretched to make the ellipse (galaxy disk outline) appear circular because this can make the bulge appear linear. A second bar score is then computed as

$$\frac{\max_i(B_{ic}^{(r)})}{\max_i(B_{ik}^{(r)})}, \tag{3.15}$$

where the indexes c and k correspond to the angles closest and perpendicular to the candidate bar angle. Finally, the bar candidate is accepted if the scores in Equations 3.14 and 3.15 both exceed empirically determined thresholds (7 and 2, respectively). As we will see in Section 3.7, when a bar is found, it is used during spiral arm-segment detection in order to help avoid bar-related false positives.

3.7 Spiral Arm-Segment Extraction

When looking for the set of arcs that describes the structure of a spiral galaxy, it may be tempting to adapt Hough transform line detection (used for bar detection in Section 3.6) for the logarithmic spiral model (see Section 3.1 and Figure 3.1 for a description of the logarithmic spiral curve). The Hough transform has been applied to shapes other than lines (see, for example, [71]), and provides parameter values for a list of detected shape instances, as we need here for logarithmic spiral arcs. We thus experimented with using the Hough transform with logarithmic spirals, but found this approach to be too problematic regardless of whether it was done in Cartesian or log-polar space (where arcs appear linear). Since the Hough transform does not naturally model or detect the endpoints of the curve used, the Hough transform does not determine all needed parameters in our case. Including endpoints in the Hough transform shape model would reduce the Hough transform to brute force search: each pixel would have to vote for a quadratic number of subsequences for each compatible combination of non-endpoint parameters.

Without endpoints in the Hough transform model, all pixels along a curve are treated the same way regardless of whether their brightness (or any other pixel property determining vote strength) is in a spatially coherent region or not. Ignoring spatial coherence disregards an important property of spiral arms, and severely reduces detection accuracy because pixels aligned with the spiral arm (regardless of distance from the arm) are considered just as strongly as actual locations within the arm. False detections can then arise from coincidental alignments of unrelated regions, while subtle spiral arm structure can be overlooked if irrelevant pixels on the extrapolated arc do not have high vote strength. Furthermore, the lengths (and thus pixel vote coverage) of lines and many other shapes (such as circles) are naturally limited by the edges of the image, but logarithmic spirals can wrap around many times within the image (or around the image sides if a log-polar representation is used). This greatly amplifies the false-alignment problem and biases vote counts toward low pitch angles (tightly wound spirals). In the extreme, a logarithmic spiral could gather Hough transform votes from all pixels in the image (once or even multiple times) by having a pitch angle close enough to zero for the distance between revolutions to be one pixel or less. Although values close to zero could be disallowed through Hough parameter discretization or limits (with the risk of information loss), this would only eliminate the most extreme cases; the bias toward low pitch angles occurs for the full range of pitch angles.

Like attempts to include endpoints in the Hough model, *post hoc* correction of these endpoint-related problems would essentially be brute-force search (possibly with some pruning heuristics); for each setting of non-endpoint parameters, all subsequences of the curve would need to be examined, and we cannot reliably use the Hough vote counts to prioritize the search because these vote counts are the values that need (potentially severe) correction in the first place. Endpoint-related problems can be partially avoided by weighting votes according to the angle agreement between the arc and the orientation at each pixel, but the angle agreement cannot be expected to be exact and the endpoint-related problems are severe, so angle weights cannot reduce false detections and fully eliminate pitch angle bias without also suppressing a significant amount of true detections. Consider, for example, how precise the orientation/arc alignment would have to be in order to suppress arcs that have multiple revolutions within an arm that is close to circular. Furthermore, although orientation-agreement weighting could narrow the set of arcs to consider, it would still be necessary to determine arc endpoints. In all, for spiral arm-segment detection, we found that the problems with the Hough transform outweighed its benefits.

Instead of using the Hough transform, we will determine spiral galaxy structure with a clustering procedure that uses orientation strengths, orientation angles, pixel proximity, and acceptable correspondence to a logarithmic spiral shape in order to group pixels belonging to the same spiral arm segment. Logarithmic spiral arcs fit to these clusters quantitatively determine the spiral structure of the galaxy. As we will see, the clustering and logarithmic spiral arc-fitting processes depend on each other: the arc fitting uses the set of pixels determined (or proposed) by the clustering algorithm, and the clustering algorithm uses the arc fit as a measure of cluster quality. We will first describe our method for fitting a logarithmic spiral arc to a pixel cluster, and then we will describe how to find pixel clusters that correspond to spiral arm segments.

3.7.1 Fitting Logarithmic Spirals to Pixel Regions

In the next section we will discuss how to use the orientation field to cluster pixels according to arm-segment membership. One of the measures of arm segment correspondence will be how well a spiral arc fits the cluster, and so we will first discuss how to fit logarithmic spiral arcs to candidate clusters. Recall that, in Section 3.1, we specified the logarithmic spiral equation as

$$\mathrm{lgsp}_{\phi,a,r_0}(\theta) = r_0 \cdot e^{-a\cdot(\theta-\phi)}, \quad 0 \leq \theta - \phi \leq \theta_e \tag{3.1 revisited}$$

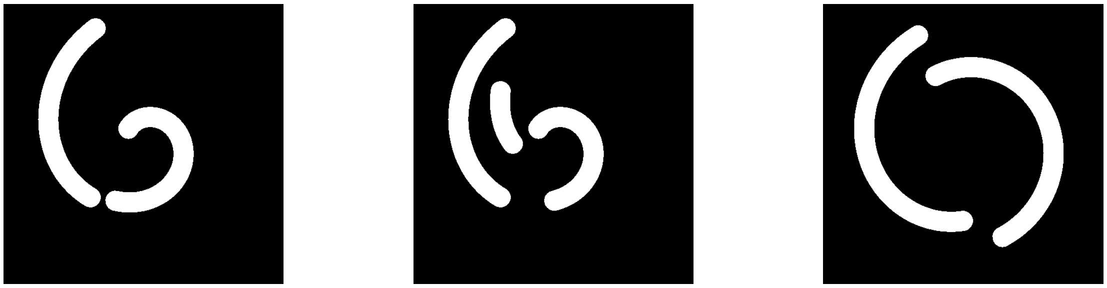
where ϕ rotates the arc about the origin, r_0 affects the size scaling of the arc, a is the tangent of the constant angle between the arc and any circle centered at the origin, and θ_e specifies the span of the arc (see also Figure 3.1). The first step in finding the parameters of this logarithmic spiral is to determine the polar coordinates of each pixel. In doing so we can use the center determined during image standardization (Section 3.3). If the cluster extends for less than one revolution about the origin and does not cross the polar axis, then we can use the standard conversion from Cartesian to polar coordinates with $\theta \in [0, 2\pi]$. If such a cluster does cross the polar axis, then we can avoid a θ-value discontinuity across the polar axis by either adding 2π to the θ values starting above the polar axis or subtracting 2π from values starting below the polar axis. Either can be done as long as the decision is consistent within the cluster; the two choices only differ in that ϕ will be exactly 2π larger in the first case, which produces an equivalent arc.

Additional care must be taken with multi-revolution clusters. Such clusters should have $\theta_e > 2\pi$ radians, but a straightforward conversion of pixel locations into polar coordinates would only have a θ range of 2π. Consequently, we only know θ values up to an integer multiple of 2π when the cluster exceeds one revolution about the origin. To determine which multiple of 2π is appropriate for each pixel, we first note that neighboring pixels should differ in θ values by less than 2π (tighter bounds are possible, but this is sufficient). Additionally, adding any multiple of 2π to all θ values does not affect the suitability of these values, so the θ value of one pixel can be fixed. Thus, for a contiguous cluster, we can first fix θ at one pixel, consider θ to be known for this pixel, and designate the rest of the pixels as unknown pixels. For any unknown pixel i adjacent to a known pixel j, we can then set $\theta_i \leftarrow \theta_i + 2\pi \cdot \text{round}(\frac{\theta_j - \theta_i}{2\pi})$ and then consider pixel j known, continuing the propagation until all pixels are known.[7]

For multiple-revolution clusters with more than one connected component, we can use a generalization of the adjacency requirement: in addition to requiring that adjacent pixels have the same θ-revolution (i.e., a difference in θ values less than 2π), it is more important for pixels to have the same θ-revolution when the pixels are nearby. Then, when expanding the set of known pixels, we always choose a known/unknown pixel pair with the shortest possible distance. Since determining θ values for non-contiguous pixels is inherently ambiguous, pathological cases are possible when prioritizing by pixel distance (see Figure 3.11). However, in such cases it is generally best to avoid fitting a single logarithmic spiral to the region in the first place (it would be less ambiguous and more descriptive to fit a separate logarithmic spiral to at least one of the connected components), and as we will see in Section 3.7.2, the logarithmic spiral fit is used to assess the quality of potential cluster merges. Considering the circumstances in which non-contiguous cluster fits will be used, a poor fit (as induced by the choice of θ values for non-contiguous cluster components) is the correct outcome because it signals that the proposed region should not be assigned a single logarithmic spiral arc.

---

[7] In practice, the pixels do not need to be processed one at a time. For example, one can often locate an angular range where only one revolution of the arc is present, split this region into two parts, compute connected components for the remaining pixels, and assign each connected component to the part it is adjacent to. This leaves only one "seam," and all pixels in one part can be adjusted at once. However, extreme caution must be taken with this type of implementation in order to insure proper handling of pixel-level variations (e.g., hook-like micro-structures) at the cluster boundary.



(a) (b) (c)

Figure 3.11: When attempting to assign polar-coordinate θ values to non-contiguous image regions covering more than one θ-revolution, prioritizing θ-consistency by pixel distance will fail in some pathological cases, but in these cases, a single cluster (spiral arm segment) should not be assigned to the image regions in the first place. Figure 3.11a is a schematic illustration of a situation where it is sensible to join non-contiguous image regions into the same cluster (though it is also reasonable to fit these regions separately). By making the nearest non-contiguous pixel pair differ in θ value by less than 2π, the fitted logarithmic spiral arc will cross the small gap (rather than a larger one). Since a single arc can fit both regions well and the gap between these regions is small, this gap was likely caused by a small dust lane or by imaging noise, rather than a difference in spiral-arm correspondence. In Figure 3.11b, the extraneous region will cause the other two regions to be joined at the "wrong" ends. However, this extraneous region belongs to a different spiral arm, so these three regions should not be grouped together in the first place. This will be correctly indicated with a poor arc fit, which discourages merges that would combine these regions. In Figure 3.11c, the two contiguous components could be a single spiral arm connected at the bottom, but they are more likely to be separate arms (and a separate fit is sensible in either case). Prioritizing the nearest non-contiguous pixel pair could choose to join either gap under small perturbations to the pixel regions, reflecting the ambiguity of the underlying situation, but in either case the combined-cluster arc fit will indicate that the two regions should be fit separately.

For a pixel cluster C, we then know polar coordinates of all pixels, and can fit the parameters of the logarithmic spiral using a nonlinear least squares minimization of the error function

$$E_C[\phi, a, r_0] = \frac{\sum_{i \in C} v_i \cdot (\rho_i - \mathrm{lgsp}_{\phi,a,r_0}(\theta_i))^2}{\sum_{i \in C} v_i} \tag{3.16}$$

where v_i is the weight for pixel i and (ρ_i, θ_i) are the polar coordinates of pixel i. For the weight values we use the image intensities after brightness transformation, but dropping the weights produces similar results. We can use a least-squares error function because we do not expect severe outliers in the pixel locations. Outliers could occur because some pixels are far away from the rest of the pixels in the cluster, or because the pixel cluster is not shaped like a logarithmic spiral (for example, it may be a combination of two arcs). In either case, the cluster does not correspond well to a single spiral arm segment. When this happens, we actually want the outlier pixels to cause a poor arc fit because we will use the fitting error to guide the selection of appropriate cluster regions (see Section 3.7.2).

In practice, we only need to fit the a (pitch angle) parameter of the logarithmic spiral (Equation 3.1). The parameter ϕ is simply the smallest θ value of any cluster pixel (or, in the case of a cluster that has a θ-span less than 2π and that crosses the polar axis, the first pixel after the θ gap). Once ϕ is determined, the optimal initial radius for a given pitch angle can be determined analytically by solving $\frac{\partial E_C[\phi,a,r_0]}{\partial r_0} = 0$ for r_0, yielding

$$\text{best\_initial\_radius}_{\rho,\theta,v,\phi}(a) = \frac{\sum\limits_{i \in C} (\rho_i v_i \operatorname{lgsp}_{\phi,a,1}(\theta_i))}{\sum\limits_{i \in C} (v_i \operatorname{lgsp}_{\phi,a,1}(\theta_i)^2)}. \tag{3.17}$$

After fixing ϕ, determining a via the least-squares fit of Equation 3.16, and calculating r_0 with Equation 3.17, the arc extent can easily be determined as $\theta_e = \max\limits_{i \in C}(\theta_i - \phi)$. Avoiding the need to fit the initial radius (and the other parameters) considerably increases the fit speed by reducing the search space. Furthermore, even though this is a nonlinear least squares fit, we have never encountered problems with local minima. The speed and reliability of the logarithmic spiral fit will be especially important because this fit will be performed frequently. Fitting not only determines the final arc parameters for each cluster, but also ensures that cluster merges do not combine multiple spiral arm segments.

A special case occurs when the cluster has an annular shape without well-defined endpoints for the spiral arc (i.e., there is no path of non-cluster pixels from the image center to an area outside the cluster). In such a situation, we know $\theta_e = 2\pi$, but the arc start angle ϕ is not well defined. For any nonzero pitch angle, the fitted arc will have a discontinuity at ϕ with a size less than the width of the cluster (if the optimal pitch angle made the discontinuity larger, the discontinuity would be large enough for the cluster endpoints to be well defined). In other words, the arc will be almost circular, with a small break. It is possible to perform a least squares fit with both ϕ and the pitch angle a as free parameters (r_0 is still calculated using Equation 3.17), but in practice we find that the discontinuity does not fit meaningful features of the cluster,[8] suggesting that these fits are noisy fluctuations around a circle. Consequently, we set $a = 0$, producing a circle since we also have $\theta_e = 2\pi$. In this case, ϕ has no impact, so we arbitrarily set $\phi = 0$. If a circle is not a good fit for the cluster, the region can be fit with more than one arc (i.e., more than one cluster, each with an angular range less than 2π), and the clustering procedure (Section 3.7.2) will decide to do so because it uses logarithmic spiral fit information in its decisions.

Since these annular clusters do not have a path of empty pixels from the image center to a point outside the cluster, we can use this criterion to detect such clusters. In practice, sometimes the cluster may have a small "crack" due to noise, such that a discontinuity in ϕ still lacks a meaningful interpretation. We can add robustness to this by applying a morphological closing to the cluster before looking for a path of empty pixels. The amount of closing should be large enough to fill cracks but small enough to avoid considering tightly wound spirals as annuli. We find it sufficient to use a radius of 3 pixels (so that gaps of width 6 pixels or less are filled), which is less than the typical cluster widths seen for the 256×256 pixel images produced in the image standardization step (Section 3.3), especially for clusters long enough to reach an angular span of 2π.

---

[8] It is not clear that annular clusters have exactly one point where the radial range somewhat sharply shifts outward. In the absence of such a pattern, the small discontinuity tends to fit noisy features of the cluster.

3.7.2 Clustering Pixels to Match Spiral Arm Segments

When looking for "good" clusters, i.e., pixel regions that correspond to spiral arm segments, we note that such clusters should be spatially coherent, should have strong and locally consistent orientations, and should encompass only one spiral arm segment. To address the spatial-coherence and orientation-based criteria, we first define a pixel-to-pixel similarity function as

$$A(v_{i_1,j_1}, v_{i_2,j_2}) = \begin{cases} \left| v_{i_1,j_1} \cdot v_{i_2,j_2} \right| & x_{i_1,j_1} \in N(v_{i_2,j_2}) \\ 0 & \text{otherwise,} \end{cases} \tag{3.18}$$

where v_{i_1,j_1} and v_{i_2,j_2} are orientation vectors for the pixels at positions (i_1, j_1) and (i_2, j_2). $N(v)$ is the 3x3 pixel neighborhood of v. By only giving nonzero similarity scores to neighboring pixels, we encode information about spatial coherence while also reducing the number of pixel similarities from quadratic to linear. The dot product gives strong values to high-strength orientation vectors with consistent orientation, encoding another criterion for good clusters. The absolute value is used because orientation direction is only unique within the range $[0, \pi)$.

To extend groupings of locally consistent strong orientation from pixel pairs to pixel clusters, we use single-link hierarchical agglomerative clustering [32]. Initially, each pixel constitutes its own cluster. The similarity between two clusters is the maximum similarity among all inter-cluster pixel pairs. Clusters merge in similarity order until the next similarity falls below a fixed threshold (empirically determined as 0.15). As illustrated in Figure 3.12, this causes clusters to "grow" within contiguous areas of strong, locally consistent orientation – properties we expect from spiral arm regions. Using single-link clustering (i.e., using the strongest inter-cluster pixel similarity as the cluster similarity) allows us to consider transitivity in locally consistent orientation: distant points on the same spiral arm generally

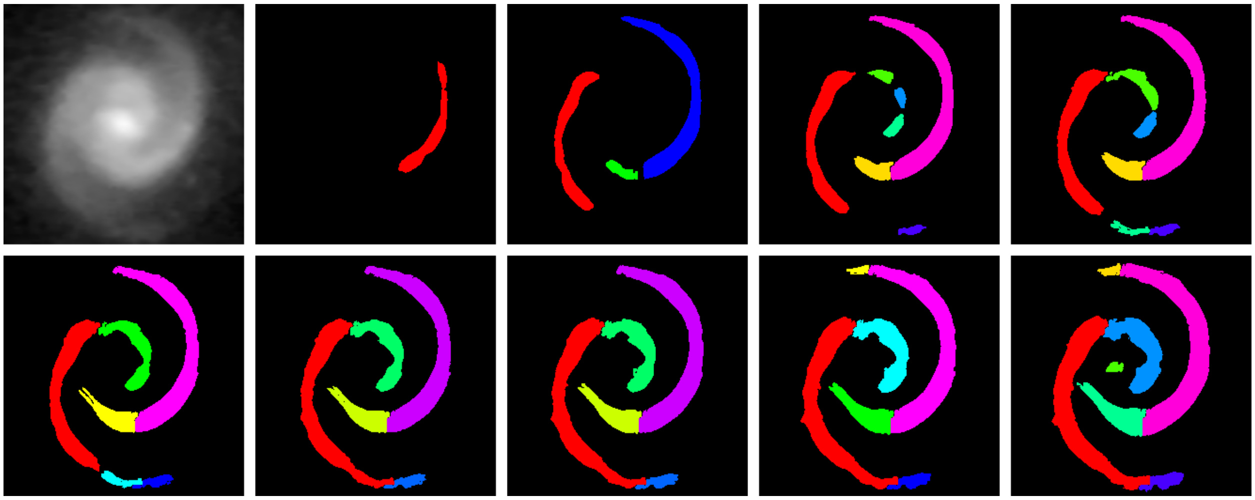


Figure 3.12: A standardized image (see Section 3.3) followed by several snapshots of the clustering process. Clusters with a size of at least 150 pixels are shown. Cluster colors can change between snapshots so that pixel-to-cluster memberships (and cluster merges) are easily visible. By merging clusters in order of the maximum-similarity inter-cluster pixel pair (with similarity defined by spatial proximity, orientation strength, and orientation alignment), the clusters grow within spiral arm segments. A secondary merge step (see Figure 3.14 and the related text) is applied afterward, so the final clusters are not shown here.

have different orientation angles due to the winding of the arm around the center, but this clustering procedure traces the chain of locally consistent pixels connecting these points. For example, in the large purple cluster in Figure 3.12, orientations near the ends of the cluster are close to horizontal and orientations near the middle are close to vertical, and these pixels are far from each other, but these pixels are nonetheless recognized as belonging to the same spiral arm segment. Furthermore, this clustering algorithm naturally determines the number of clusters in the image, so we do not have to make assumptions about the spiral arm-segment count, or attempt to determine it separately.

When clusters are built purely from pixel-level orientation field similarities, the clusters do tend to follow spiral arms, but may combine regions that would best be modeled separately (e.g., arm forks, or bent arms). Enforcing this third property of good clusters, correspondence with only one spiral arm segment, can be done using logarithmic spiral arc fit quality. We note that clusters with poor arc fits can arise once clusters grow large enough for the chain of


(a) (b) (c)

Figure 3.13: Incorporating logarithmic spiral fit information into the clustering guards against fitting two spiral arm segments with one logarithmic spiral arc. When two spiral arms are adjacent to each other, the orientation vectors can flow smoothly throughout the junction, rather than defining a clear boundary between spiral arm segments. This is shown (at half resolution for display purposes) in Figure 3.13a. Thus, clustering based solely on orientation vectors can combine two arm segments into the same cluster. An arc fit to such a combined cluster does not describe the two arm segments well (Figure 3.13b). If merges are blocked when the increase in fitting error is too high (see text), the two arm segments are assigned separate clusters, greatly improving description accuracy (Figure 3.13c). With the fork shown here, it is ambiguous which cluster the lower region belongs to, but both interpretations are reasonable, and they both describe the fork structure. The ambiguity is naturally decided via orientation vector similarities (Equation 3.18) because these similarities define the cluster merge order. Since the right prong of the fork has higher orientation vector similarities at the junction point, the lower region will be combined with the right prong of the fork.

locally consistent orientations to potentially branch from one arm segment into an adjacent arm segment (see Figure 3.13a). This can be prevented by blocking cluster merges where the individual clusters have much better logarithmic spiral arc fits than the combined cluster. When two clusters C_1 and C_2 are about to merge, we fit a logarithmic spiral arc to the combined cluster, and assess the error increase for each constituent cluster. In particular, we use Equation 3.16 to calculate the merge-badness score

$$G[C_1, C_2] = \max\left(\frac{E_{C_1}[\phi^{(m)}, a^{(m)}, r_0^{(m)}]}{E_{C_1}[\phi^{(1)}, a^{(1)}, r_0^{(1)}]}, \frac{E_{C_2}[\phi^{(m)}, a^{(m)}, r_0^{(m)}]}{E_{C_2}[\phi^{(2)}, a^{(2)}, r_0^{(2)}]}\right) \tag{3.19}$$

where $(\phi^{(i)}, a^{(i)}, r_0^{(i)})$ is obtained from the least-squares fit to C_i as described in Section

3.7.1. For $(\phi^{(m)}, a^{(m)}, r_0^{(m)})$ we perform a least-squares fit to the merged cluster $C_1 \cup C_2$, with the weights v_i in Equation 3.16 re-scaled so that $\sum\limits_{i \in C_1} v_i = \sum\limits_{j \in C_2} v_j$. If $G[C_1, C_2]$ exceeds an empirically determined threshold of 2.5, the combined fit is unlikely to properly model at least one of the two clusters, so the merge is blocked. If the merge between C_1 and C_2 was blocked, it is tried again later if a subsequent inter-cluster pixel pair also joins these clusters, since one or both clusters may have grown (changed shape) in the interim. We find that the threshold of 2.5 is permissive enough to allow relatively good merged-cluster (single-arc) fits even though better fit errors can always be obtained with two arcs (and thus more free parameters), while remaining strict enough to avoid merging two regions that should be modeled separately (with one possible exception discussed and handled below). Figure 3.13 illustrates how this fit-based merge check can avoid merging two different spiral arm segments into the same cluster.

We only check this condition when both clusters have reached a minimum size, set to $\frac{R}{10}$, where R is the image resolution in the shorter dimension. This saves some computation time, reflects the fact that clusters can only encompass multiple arm-segment regions once the clusters are sufficiently large, and avoids fitting until a reasonable cluster shape can be determined. Also, we can optionally refine the merge-badness score to avoid distortions from small "bumps" or other jagged features of the cluster boundary. These small jagged areas can disproportionately increase the fit error of both the individual and merged clusters. This increases both the numerator and denominator of the merge-badness score in Equation 3.19, so that jagged boundaries can make the combined cluster fit appear better than it would be under the intended criterion of logarithmic spiral arc compatibility alone. This effect can be counteracted by computing squared fit error values for each pixel in each (pre-merge) cluster, temporarily ignoring pixels that have squared error z-scores (number of standard deviations above the mean) with absolute value greater than 3, and recomputing the merge-badness score in Equation 3.19 using the remaining pixels. The cutoff of 3 standard deviations is a conservative threshold that only ignores a small number of pixels (if any), so that small jagged areas can be ignored, but larger deviations (arising from the arc shape of the cluster as a whole) are still considered. If an even more permissive threshold is used, even fewer pixels are ignored, and the jagged-region removal effect becomes negligible or nonexistent. If a more aggressive threshold is used, the merge-badness measure becomes too reliant on exact correspondence to a logarithmic spiral arc. In practice, we find that ignoring jagged regions is only a small refinement (and can usually be omitted without consequence). When two clusters have incompatible logarithmic spiral arcs, the effect on a single-arc fit is generally much stronger than the influence of small jagged parts of a cluster (especially since, as mentioned in Section 3.7.1, the squared error is much more sensitive to severe outlier values that result when an entire cluster deviates from a single logarithmic spiral arc, as compared to pixels that are slightly more distant than normal).

Clusters also grow within the bar region (if the galaxy has a bar), so we must also account for the possibility that the cluster corresponds to a bar instead of an arm segment. To do so we use our earlier determination of bar presence and location (Section 3.6). The bar-fitting error B_C is computed as the weighted mean squared distance from the line segment defined by the bar angle and radius, when a bar detection was made earlier (otherwise, $B_C = \infty$). The weights (if any) are the same ones that would be used in the logarithmic spiral arc fit, making the bar fit error directly comparable to the logarithmic spiral arc fit error. If $B_C < E_C[\phi, a, r_0]$, we consider the cluster as part of the bar and substitute B_C for $E_C[\phi, a, r_0]$ in Equation 3.19. Bar clusters can then grow in the same way as spiral arm segment clusters, without the risk of considering the bar as an arm, and with a more appropriate model for bar clusters during merge checks.

In cases where the galaxy was highly inclined relative to our line of sight, the bulge can appear linear after correcting the disk to appear face on (see Section 3.3 for a description of this process). Since the bulge can appear linear, clusters can also grow in the bulge region, so after the clustering described above we delete the cluster containing the center pixel (if such a cluster exists and is not already described by a bar). We find that such clusters rarely contain spiral arms; this is likely because a merge with the center-containing cluster requires a multi-revolution spiral at the center, and since this spiral must fit the arm along with the bulge, the fit quality will be poor for at least part of the bulge region. This poor fit quality then discourages merges via larger values of the merge-badness score (Equation 3.19). Nevertheless, it would be ideal to have a bulge model used in the same way as the bar model (i.e., replacing the logarithmic spiral arc with the bulge model if the latter is a better fit) as additional discouragement against encompassing the bulge and an arm in the same cluster. Even with center-containing cluster deletion, bar detection is still necessary because galaxies that are not highly inclined can have an orientation-field gap at the bulge region (image center), so bar clusters may not contain the center in this case. Additionally, using the bar model in fit-based merge checking discourages arms from merging with the bar region, which also discourages arms from further growing into the bulge region (image center).

By blocking proposed merges that yield poor fit quality, we ensure that we do not try to describe multiple spiral arms (or an arm and a bar) with one logarithmic spiral arc, greatly improving the accuracy of our structure description. This merge checking also reduces sensitivity to the stopping threshold used in the hierarchical agglomerative clustering; we can set this value low enough to allow as much meaningful growth (and merging) of clusters as possible, without as much concern about unwanted merges. Another benefit is that we can incorporate arc-shape knowledge into the clustering without making the logarithmic-spiral assumption too strong. Arms that deviate from the logarithmic spiral model are naturally split into multiple logarithmic spiral segments that can, when needed, form a piecewise approximation to the actual curve of the arm. If a user wishes to perform a fit with a different model, the arcs' associated pixel regions can be used. Additionally, clusters are formed using pixel-level information (following the "flow" of high-strength orientation vectors) rather than explicitly searching for logarithmic spiral curves; the logarithmic spiral fit can only decide which merges of pre-existing clusters are close enough to a logarithmic spiral.

This fit-based merge checking is thus beneficial overall, but it can cause slight over-fragmentation in the clusters. Since merge blocking occurs while the clusters are still forming, it is possible for a pair of final clusters to be compatible with a single arc, even if earlier stages of these clusters were not. Using a higher (more permissive) threshold in Equation 3.19 is not feasible because it risks allowing merges of adjacent spiral arms, which is much worse for description quality than using two arcs to describe one spiral arm. It is possible to periodically re-check merges that were previously blocked (once either or both of the two clusters have grown), but this would substantially increase computation time, especially because many clusters remain unsuitable for merging, for a prolonged period or permanently.

Another source of cluster over-fragmentation is obscuring elements (such as dust lanes and brightness variations) that create small gaps in the orientation field. Such gaps could be bridged by increasing the neighborhood size in Equation 3.18, but this would drastically increase the number of nonzero pixel similarities to consider. Even if a sufficiently large neighborhood size still yielded reasonable computation time, the ability to jump over pixel gaps would often need to be suppressed in order to avoid losing spatial coherence in the clusters, and brightness gaps often indicate genuine separations between spiral arms. Gap-crossing is thus undesirable until we can restrict it to the small fraction of cases where there is good reason to ignore spatial discontinuities.

Both sources of cluster over-fragmentation can be addressed by applying a second, purely fit-based merge step after the earlier clustering has finished. At this point, the number of clusters is small enough to quickly consider all potential merges of nearby clusters. Additionally, the close-to-final arc shape of the clusters is known (since most merges have already occurred), and to address any remaining uncertainty, we can merge the most-certain cases first by

Figure 3.14: Pixel clusters before (left) and after (right) a secondary merging step performed after the main pixel clustering. The secondary merging step joins clusters that are compatible with the same logarithmic spiral arc, but were not previously merged because logarithmic-spiral compatibility was not clear in an earlier stage of the clustering (before near-final cluster shapes were known), or because there was a small gap in the orientation field (these gaps are usually meaningful, but can be jumped over if justified by sufficient compatibility in the near-final logarithmic spiral arc fit). Cluster colors differ between images so that pixel-to-cluster memberships (and cluster merges) are easily visible.

merging in order of combined-cluster fit quality. Knowing the larger-scale arc shape of the clusters also means we have have enough information to justify bridging small gaps in brightness.[9]

In this secondary merge (i.e., defragmentation) step, we calculate merge scores $G[C_1, C_2]$ (Equation 3.19, including the substitution of the bar score if it is better) for all cluster pairs with maximum Euclidean distance $\frac{R}{20}$ between the closest inter-cluster pixel pair. These clusters are then merged in score order until no pair's merge score remains below the previously mentioned threshold of 2.5. If a merge is performed, all scores involving the newly merged cluster are updated. Using the distance threshold (instead of requiring adjacency) allows the merging to jump over reasonably small brightness gaps if justified by the fit quality of the

---

[9] We did consider methods for bridging larger brightness gaps, as motivated by a few (atypical) cases where one cluster visually appeared to be along the arc of another cluster despite a large gap between clusters. However, crossing large brightness gaps requires extrapolation over a long distance and increases reliance on the assumption that clusters have a logarithmic spiral shape. These considerations, combined with the fact that large brightness gaps are often even more meaningful than smaller ones (and thus less likely to be a false separation), plus the risk of false alignments, meant that long-distance merging was too unreliable. Instead, we found that "true" long-distance merges were better achieved by improving the brightness transformation (Section 3.2) to make the orientation filtering as sensitive to arms as possible, which changed large gaps to small gaps in many cases. When improvements to the orientation-filter sensitivity did not make a difference, we leave the clusters as is, because such a gap is likely to be informative, and users can re-examine separate clusters more easily than combined clusters.

merged cluster. Considering all pairs of nearby clusters ensures that all previously-blocked merges are reconsidered with the near-final cluster shapes. Merging in score order (and thus merge-certainty order) addresses any remaining tentativeness in the cluster shapes. Figure 3.14 illustrates how this secondary merging step combines clusters corresponding to the same spiral arm.

After the cluster merges have been finalized, we report clusters (and their corresponding logarithmic spiral arcs) that are sufficiently large. For 256 × 256 images, we require that clusters have a size of at least 150 pixels. This size cutoff poses a tradeoff between arc-detection sensitivity and noise resistance. Since it is much easier for a user of our output to impose a stricter cutoff (versus a more permissive cutoff) on their own, and since clusters can be weighted by size (or related measures, such as logarithmic spiral arc length), we use a fairly permissive cutoff.

At the end, we have a set of pixel clusters, each with a corresponding logarithmic-spiral or bar parameterization. This information can be used as-is as a description of general spiral galaxy structure (for comparison with simulations, for example), or it can be used to calculate galaxy-level quantities such as spiral arm winding direction, tightness, asymmetry, or other measures of interest. Our structure-extraction method does not require a brute-force search for all possible logarithmic spiral arcs, but rather uses the orientation field and spatial proximity to build image regions corresponding to spiral arms. Higher-level information, namely the expectation that clusters should be shaped like a single arc (or a linear bar), is incorporated without being too stringent in the requirement for logarithmic spiral shape; the clusters can still grow along the arbitrary shapes defined by the orientation field, and the pixel-to-cluster assignments are available along with the corresponding logarithmic spiral arcs. This procedure can thus produce a description of arbitrary spiral arm-segment structure, and does so without human supervision. In Chapter 4 we will detail the information available from this method, and in Chapter 5, we will evaluate the accuracy of this method by comparing it with human determinations of spiral galaxy structure.

Chapter 4

Information Available From Our Method

As discussed in Section 3.1, we represent spiral galaxy structure as a list of logarithmic spiral arcs associated with pixel clusters, such that each parameterized arc (and pixel cluster) corresponds to one spiral arm segment. This information can be used to calculate measures of spiral galaxy structure; for example, we have provided measures of winding direction, arm tightness, and arm count in our output, as discussed later in this chapter as well as in Chapter 5. These measures are likely useful in their own right, and we also expect that our list-of-arcs (and clusters) output is precise and general enough to facilitate new measures and astronomical applications beyond those discussed or even anticipated in this work. To facilitate these uses, we provide a detailed description of the information available from our method.

Our method provides spiral galaxy structure information through three types of outputs: a set of images with information about various steps of our method, a CSV file that provides arc-level (cluster-level) information, and a CSV file that provides galaxy-level information.

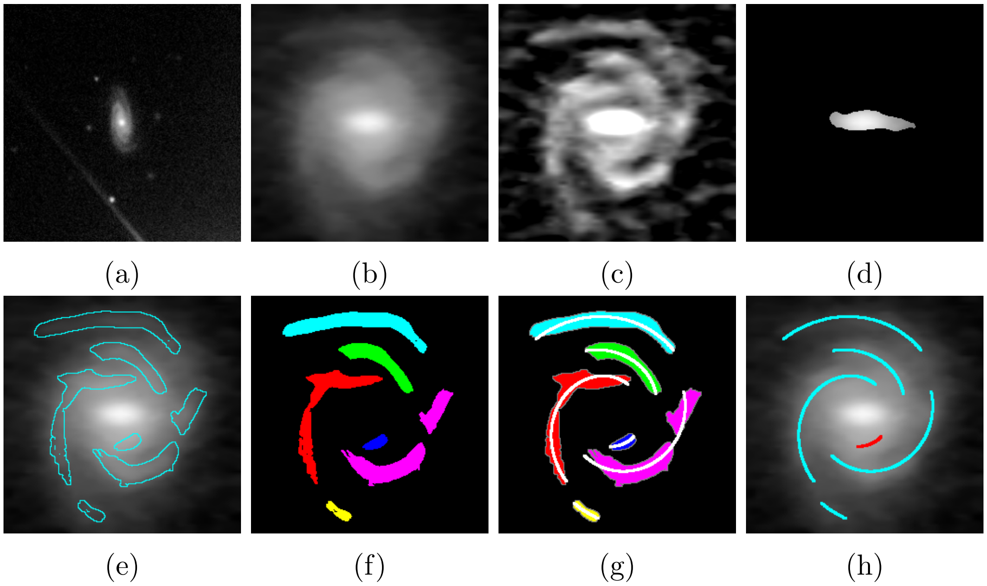

(a) (b) (c) (d)
(e) (f) (g) (h)

Figure 4.1: Available image outputs from our method: (a) the input image, (b) the image after image standardization (Section 3.3), (c) the standardized image after unsharp masking (Section 3.5.2), (d) the center-containing cluster deleted from the final output (if such a cluster exists; see Section 3.7.2), (e) the standardized image overlaid with the final cluster boundaries (including the bar cluster, if any, but this image does not have a bar cluster), (f) the final pixel-to-cluster assignments, with each cluster given a unique RGB color, (g) the clusters overlaid with the final logarithmic spiral arcs associated with each cluster, and (h) the standardized image overlaid with these arcs. The arc-fitting and clustering procedures are discussed in Sections 3.7.1 and 3.7.2, respectively.

All of the arc-level information is provided in terms of the standardized (de-projected) image, but the galaxy-level CSV provides the detail needed to re-project any galaxy- or arc-level information.

4.1 Output Images

The set of images provided by our method is displayed (for an example galaxy) in Figure 4.1. This image set consists of: the original input image; the image after image standardization (Section 3.3) and then after unsharp masking (Section 3.5.2); the deleted cluster (if any; see Section 3.7.2); the boundaries of the pixel clusters overlaid on the standardized image (including both spiral arm-segment and bar clusters); the exact pixel cluster regions, where each cluster is assigned a different RGB color; the pixel clusters with the fitted spiral arcs (and fitted bar, if it exists) overlaid; and the standardized image with the fitted spiral arcs (and fitted bar, if it exists) overlaid. A PDF showing the orientation field could also be provided, but we do not produce it by default because it requires much more time to create (about 3-4 seconds) than the other images, which may be significant for large image sets. In the final image (the standardized image with the arc and bar overlay), arcs with a S-wise winding direction are shown in red, arcs with a Z-wise winding direction are shown in blue (cyan in Figure 4.1h for easier visibility at a smaller image size), arcs with neither a S-wise nor a Z-wise winding direction (i.e., arcs with pitch angle zero) are shown in magenta, and the bar (if present) is shown in green. Most of these images are primarily useful for visualizing the general idea of what our method "saw" in a galaxy, but the cluster-assignment image (Figure 4.1f) also provides the set of pixels associated with each logarithmic spiral arc (or bar). These pixel assignments could be used to perform other measurements of the corresponding spiral arm segment (or bar), such as a measure of the brightness distribution.

4.2 Arc-Level Information

The per-arc CSV file lists each galaxy multiple times: once for each arc found in the galaxy image. The first column gives the galaxy name (an echo of the image name). For each galaxy, each arc is listed in order of arc length; this arc-length rank is provided in the second column. Next are the core logarithmic-spiral parameters given for each arc (see Section 3.1 and Figure 3.1). These parameters are expressed in terms of the standardized image and use the estimated (fitted) galaxy center given in the galaxy-level CSV (see Section 4.3). The pitch angle gives the constant angle between the arc and any circle centered at the estimated galaxy center; the sign of the pitch angle gives the winding direction of the arc. The initial angle (the ϕ parameter) gives the counterclockwise angle from the polar axis (in the image, the theoretical line segment extending horizontally from the estimated galaxy center to the right side of the image) to the start of the arc. The initial radius is the distance (in pixels within the standardized image) from the estimated galaxy center to the arc at this starting angle value. This distance may be the closest or farthest point on the arc, depending on the winding direction of the arc (sign of the pitch angle). The angular extent gives the unsigned angle subtended by the arc; adding this value to the initial angle gives the ending angle of the arc. The cluster output color gives the RGB color associated with the cluster in the cluster-assignment image (Figure 4.1f), enabling measurements of brightness, cluster area, or other properties.

The per-arc CSV file also provides several values that could be computed from the arc parameters and/or pixel cluster membership information, but are provided for convenience. The `rStart` and `rEnd` variables give the smallest and largest distances (in standardized image pixels) from the estimated galaxy center to the arc. One of these values is equivalent to the initial radius (`rStart` for Z-wise arcs and `rEnd` for S-wise arcs); the other can be computed from the logarithmic spiral function (Equation 3.1) at the end of the arc (initial angle plus angular extent). The `numPixels` and `arcLength` columns give the size of the cluster and the length of its arc, in units of standardized image pixels. We then give two values for the cluster's weighted squared fit error (Equation 3.16): one normalized by arc length and the other normalized by the number of pixels (we are not sure which is more astronomically meaningful, or if they both are to various extents in different contexts, so we provide both). The `meanIntensity` column gives the average cluster brightness; this brightness is after brightness transformation (Section 3.2), so while this value gives some information about the visual distinctiveness of the cluster, its value is unlikely to be astronomically meaningful. If it is desirable to determine the average of the original brightness values within a cluster, the FITS brightness values could be de-projected, or the cluster membership image could be re-projected, based on the values given in the per-galaxy CSV (Section 4.3).

Finally, the arc-level CSV file provides several flags about the state of the fitting procedure used to determine the logarithmic spiral parameters of the clusters. The `used2rev` flag indicates whether the angle subtended by the cluster exceeded 2π radians; this requires additional fitting steps (see Section 3.7.1). Such a condition is not problematic unless the `failed2rev` flag is set, which indicates that our code was unable to disambiguate θ values for the cluster (Section 3.7.1). Such a failure is reasonable when evaluating the potential merge of two long spiral arms that should not be merged together, but this failure should not occur in our final output, and we are not aware of a case of this happening in the images we tested. The `hasUndefinedEndpoints` flag indicates whether we encountered the annular-cluster scenario described in Section 3.7.1; in this case, the pitch angle is fixed to zero, the angular extent is 2π radians, and the initial angle is not meaningful (it is arbitrarily set to zero).

4.3 Galaxy-Level Information

The galaxy-level CSV file provides the status and any warnings regarding the fit to each galaxy, the parameters used to standardize (de-project) each galaxy image, and (for convenience) galaxy-level structure measurements. The `fitState` column is a text field that indicates whether the structure-extraction procedure could be completed for the image (in which case the string “OK” is given), or whether there was a problem severe enough to necessitate skipping the image (in which case an error description is given). Currently, the only problem encountered occurs when no clusters exceed the minimum size limit (and, as will be seen in Chapter 5, this is very rare for spiral galaxy inputs), but if other errors arise, they would be given in this column. The `warnings` column can provide zero or more strings indicating that the image could be difficult to fit, or that an unusual condition was encountered during the fitting. Indicators that the image could be difficult to fit include a low detected galaxy resolution as measured by the minor axis length of the ellipse used for de-projection (Section 3.3), or a large drift in the measured galaxy center when a star mask is not available. Unusual conditions encountered during fitting include the need to delete a center-containing cluster (Section 3.7.2) or the presence of an arc that has unusually high error per unit length. These warnings do not necessarily indicate that something is wrong with the output, and in most cases these warnings can safely be ignored. However, if one wishes to manually review some images in a large set, these warnings may give some indication of where it may be most useful to look for potentially-problematic cases.

The `starMaskUsed` and `noiseMaskUsed` columns indicate whether a star mask or noise mask was used, and if so, the level of aggressiveness needed (see Section 3.4). The possible values can be "unavailable" if no mask was provided, "none" if the mask was provided but not used, or "conservative," "aggressive," "last-resort," or "fail" for the various levels of aggressiveness. If the star or noise mask is listed as "fail," fitting may still succeed on the image (for example, the fitted galaxy center may legitimately be somewhat distant from the image center), but there is a higher risk of star-related problems. The `hasDeletedCtrClus` flag indicates whether a cluster was deleted because it contained the center pixel but the bar candidate, if present, was not a better fit than a logarithmic spiral arc. The `failed2revDuringMergeCheck`, `failed2revDuringSecondaryMerging`, and `failed2revInOutput` flags indicate whether our code was unable to disambiguate θ values for a multi-revolution cluster (Section 3.7.1). As mentioned in Section 4.2, such a condition is usually not problematic if it occurs during a merge check or during the secondary merging (some proposed merges are so unreasonable in terms of a combined-cluster single-arc fit that there is not even a sensible multi-revolution θ disambiguation); it is only a concern if the `failed2revInOutput` flag is set, since this would mean that one of the output clusters may have been fitted improperly. We have never encountered this in the images used anywhere in this work, but if it does arise, one can examine the arc-level CSV file to see which arc was affected.

The galaxy-level CSV also provides information about the fitted ellipse used in image standardization (Section 3.3) so that the de-projection operation can be replicated for other images (for example, the FITS image giving the original brightness values) or applied in reverse to our output images (e.g., to map our detected clusters to regions in the original FITS image). The main CSV values to use for this purpose are `inputCenterR`, `inputCenterC`, `diskMajorAxisAngleRadians`, `diskAxisRatio`, and `diskMajorAxisLength`. All of these values are given in terms of the input image (not the standardized image). The center values `inputCenterR` and `inputCenterC` specify the row and column of the center, with sub-pixel precision and with the pixel at the upper left of the image corresponding to row 1, column 1.[1] The value `diskMajorAxisAngleRadians` gives the counterclockwise angle from the polar axis (the imaginary horizontal line from the estimated galaxy center to the right edge of the image) to the major axis of the ellipse, `diskAxisRatio` gives the ratio of the minor axis length to the major axis length, and `diskMajorAxisLength` is the length of the major axis, in units of input image pixels. Together, these values provide the position and shape of the fitted elliptical outline of the galaxy disk. As discussed in Section 3.3, the image is rotated about `inputCenter` so that `diskMajorAxisAngleRadians` becomes 90 degrees, stretched according to `diskAxisRatio` so that the disk outline becomes circular, and then cropped so that the bounding square has length `diskMajorAxisLength` on each side. For convenience, the CSV also provides `standardizedCenterR` and `standardizedCenterC` (the exact, sub-pixel location of the fitted ellipse center after image standardization) as well as `diskMinorAxisLength` (which can be useful for identifying galaxy images with low input resolution, or images where the ellipse fit may have zoomed in too far).

We also provide information about our bar detection, though this should be viewed mostly in terms of understanding the state of our fit procedure for a particular galaxy, rather than as a way to obtain reliable bar detections (as discussed in Section 3.6, bar detections are only relevant to our goals when the bar may otherwise be mistaken for a spiral

---

[1] This follows the convention in MATLAB, where the code was originally implemented.

arm). The column `barCandidateAvailable` indicates whether a bar's two detection scores exceeded their thresholds (discussed in Section 3.6; the scores are given in CSV columns `barCandScoreOrifld` for the orientation-field-based score and `barScoreImg` for the score based on transformed image brightness values). If a bar candidate is available, then the fit error of the bar (under fixed bar parameters) was used in merge decisions where appropriate (see Section 3.7.1). If, at the point of producing arc-level output, arc parameters were not given for a cluster because the bar was a better fit, the `barUsed` column will be set to true. Some clusters may both contain the center pixel and have a better fit to the bar, in which case the bar substitution takes precedence over cluster deletion. The bar parameters (radius and angle) are given as `barHalfLengthStandardizedImage` and `barAngleStandardizedImage`. These values are given regardless of whether a bar candidate is available (our method tries to find the most plausible bar parameters for every image, regardless of whether the bar actually exists), but should only be considered meaningful if the `barCandidateAvailable` column (and perhaps also the `barUsed` column) is set to true.

Next, the galaxy-level CSV provides a few basic statistics about the distribution of arc lengths in the image, particularly in the degree of fragmentation. The columns `totalNumArcs` and `totalArcLength` simply give the number of arcs and their total length. Basic information about the distribution of arc lengths is given as the average, median, minimum, maximum, and upper and lower quartile arc lengths. Arc fragmentation is measured more directly by considering the arcs laid end to end and sorted in decreasing order by length. We can then consider the length and rank of the arc at 25, 50, and 75 percent of the total arc length (columns `alenAt50percent` and `rankAt50percent`, and defined analogously for the other percentage values). These can be interpreted as arc-length-weighted median and arc-length-weighted quartile values. Galaxies with longer arcs at a given percentile (and perhaps also with higher concentrations of arc length) may have better-defined arm segments, although many galaxies inherently have fragmented arms, so in these cases fragmented arcs are an inevitable consequence of a proper structure description. If a user is interested in other measures of arc fragmentation, such measures can be computed from the arc-level CSV described in Section 4.2.

For convenience, the galaxy-level CSV also provides several galaxy structure measurements (but users may use the arc-level CSV to compute measurements of other structural properties, or calculate existing structure measurements in alternate ways). For winding direction, the CSV provides S-wise and Z-wise classifications according to a simple majority vote of the arcs (`chiralityMaj`), the winding direction of the longest arc (`chiralityLongest`), or the arc-length-weighted vote of all arcs (`chiralityAlenWtd`). The underlying vote counts are also given with one vote per arc (`chiralityVotesMaj`), and for the arc-length-weighted vote (`chiralityVotesAlenWtd`). Another column, `top2chiralityAgreement`, indicates whether the two longest arcs are reasonably long (64 pixels, i.e., a quarter of the number of pixels along one dimension of the standardized image) and agree in winding direction. If both of the two longest arcs meet this length threshold, the value of this column will be "agree" or "disagree." If there is only one long arc, the value will be "one-long," and if there are no long arcs, the value will be "all-short" unless there are less than two total arcs (in which case the value is "$<$ 2 arcs").

The pitch angle (arm tightness) of a galaxy depends on the pitch angles of its arms, which often vary within a galaxy. Consequently, our CSV reports several possible galaxy-level pitch angle measures: the pitch angle of the longest arc alone; the average pitch angle of all arcs, regardless of length or other properties; average pitch angles weighted by length or brightness; and finally, a length-weighted average based only on arcs with winding direction agreeing with the arc-length-weighted dominant winding direction. We believe that the latter measure may be the best candidate for "the" pitch angle of a galaxy, if one number is to be assigned to the pitch angle of a galaxy, and we use this measure when comparing to human classifications (see Section 5.2). However, the user is free to choose whichever measure of pitch angle they wish, or compute their own from our arc-level information. In addition to the pitch angle, we also provide a measure of the error in this pitch angle (considering only the spread of pitch angles detected in the galaxy): an arc-length weighted standard deviation of the arcs that agree with the arc-length-weighted dominant winding direction.

Although the notion of arm count is ambiguous (since 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. Such counts are likely a useful starting point, but we encourage users to calculate arm count (using the arc-level CSV file) in a way that best suits their needs.

4.4 Uses of Our Output

Our galaxy-level CSV file provides several measures of fragmentation, winding direction, pitch angle, and arm count, but we intend these measures to only be a starting point. Plenty of other measures of galaxy structure (e.g., lopsidedness, or the presence or absence of a ring) could be computed from our output, and existing structural measures could be computed in more sophisticated ways (possibilities include incorporating brightness information, excluding or down-weighting arcs with a high error per length, and calculating pitch angle using only the arcs at least partially within a certain radial range within the galaxy disk). Furthermore, the choice of some measures may depend on the application (e.g., what one considers to be an arm, or the kind of arm one is interested in, when computing arm counts). In general, our intent is to provide information from which application-appropriate measures of spiral galaxy structure can be readily calculated.

In addition to determining properties of a given galaxy, our method can also be used to

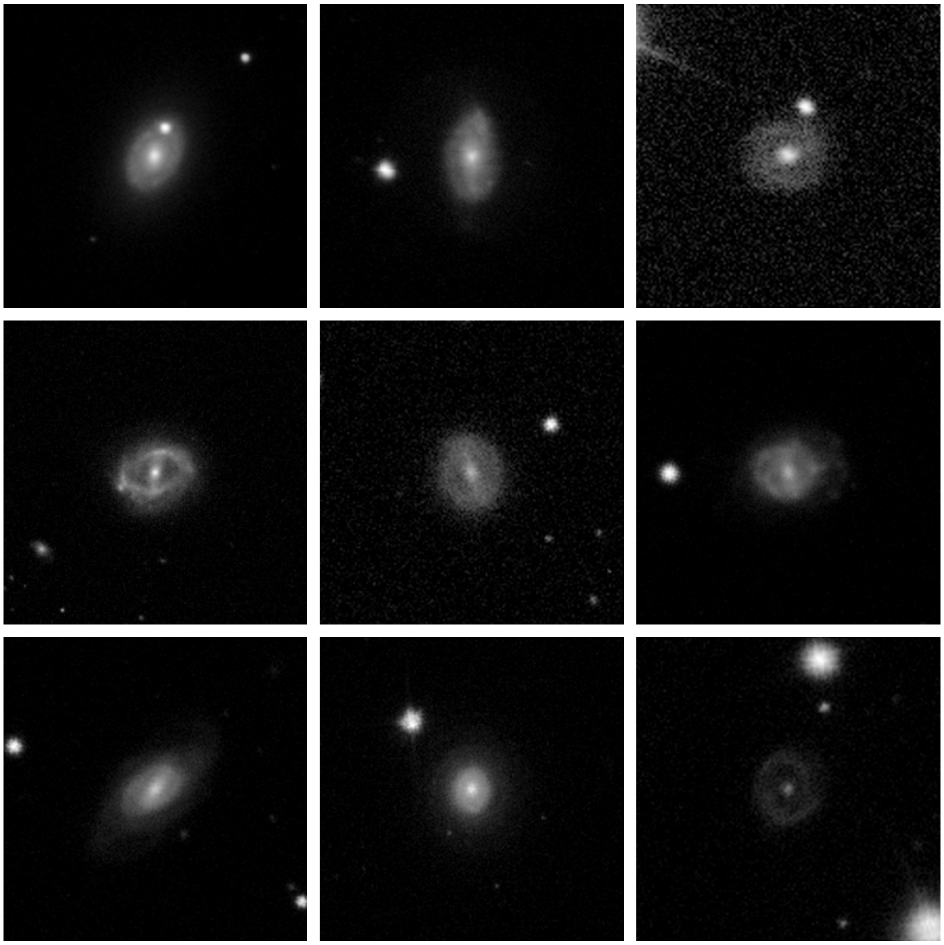


Figure 4.2: Our output can be used to produce an arc database that can be queried to find galaxies with the desired morphological criteria. Here we look for galaxies with just one low-pitch-angle arm segment, yielding typical examples of “ring” galaxies.

find galaxies with specific properties. For example, our output can be used to analyze distributions of galaxy measurements in order to find instances of “weird” morphologies, and can also be used to organize the detected-arc information into a database that can be queried to return galaxies with user-specified 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 4.2 depicts nine typical objects found using this query.

Chapter 5

Comparisons With Other Spiral Structure Information

Figure 5.1 provides several visualizations of our output when run on galaxies where fairly clean images are available. As we can see, our output is also fairly clean, tracing the arms in a way that would be plausible to a human. As image quality degrades, we expect both human and automated designations of spiral structure to be affected, but we find that we still attain good agreement with humans (where and to the extent that agreement can be determined), as we explore next.

To the best of our knowledge, this work provides the first means of automatically determining

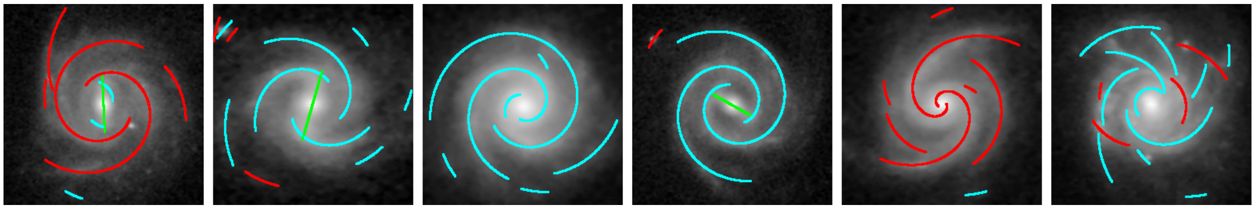


Figure 5.1: Typical examples of how our method performs on galaxies where fairly clean images are available. Green represents a bar; red arms wind S-wise, and cyan arms wind Z-wise.

arbitrary arm-segment structure in spiral galaxies. We are not aware of any previous large-scale catalogs providing this information, either manually or with an automated method. As such, we do not have a basis for directly comparing our output with other methods or studies. We can, however, compare properties computed from our output, namely spiral arm winding direction, spiral arm tightness, and (to some extent) the number of spiral arms. These values can be computed from our output in many ways (including measures more complicated than those considered in this work, which may have even better correspondence to human classifications), and the spiral galaxy properties under comparison do not cover the full range of information that can be determined from our output. Nevertheless, comparisons can provide information about the accuracy of our method (to the extent that the properties are well-defined and the previous structure property measurements are accurate). Furthermore, the measures calculated from our method provide case studies in how structural properties of interest can be calculated from our output, and may provide guidance in determining properties (such as the degree of asymmetry or flocculence) not included here.

As mentioned in Section 2.6, Galaxy Zoo [40, 38] and Galaxy Zoo 2 [67] provide information about spiral arm winding direction, tightness, and count through classifications provided by hundreds of thousands of human volunteers. For our comparisons with each of these classifications, we use a sample of 29,250 galaxies where the human classifiers indicated visible spiral structure, as chosen by the director of the GZ2 project (Steven Bamford, personal communication). For these galaxies we apply our code to red-band FITS images from the Sloan Digital Sky Survey [69], processed as described in Chapter 3. Some other studies specifically measure winding direction or pitch angle, so we compare our results with these measures as well.

5.1 Spiral Arm Winding Direction

Spiral arm winding direction is the classification of a galaxy's spiral arm winding pattern as clockwise or counter-clockwise. Following Willet et al. [67] and others, we refer to these categories as "Z-wise" and "S-wise" to avoid ambiguity (otherwise, for example, differing winding direction conventions could be followed depending on whether one visually traces the arcs in an inward or outward direction). This naming also emphasizes that the classification is based on visual information.[1] Winding direction is a simple property of spiral arm structure, but it is nonetheless useful in astronomy and cosmology (see Section 1.2). Furthermore, winding direction has a well-defined answer for many (but not all) spiral galaxies, allowing the comparisons to be conducted with less ambiguity than other aspects of spiral structure.

In Galaxy Zoo 1 (GZ1), volunteers chose one of six categories for each galaxy, including two for winding direction: Z-wise spiral galaxy and S-wise spiral galaxy [40, 38]. The other four categories are "edge on/unclear," "elliptical," "star/don't know," and "merger." These categories and their corresponding icons (presented to the human classifiers) are shown in Figure 5.2. Although the galaxies used in our comparison were selected for spiral feature visibility, galaxies still vary in winding direction clarity. Consequently, human classifiers viewing the same galaxy may agree or disagree with each other on the winding direction, and on whether a winding direction is visible at all. For each galaxy, we thus calculate a "discernibility" score as

$$\frac{\max(\text{S-wise votes},\ \text{Z-wise votes})}{\text{total number of votes}}.$$

For example, a discernibility of 60% indicates that 60% of humans voted for one winding direction, while the other 40% voted either for the other winding direction, or for edge-on (i.e.,

---

[1] We are extracting information about the spiral pattern, not the rotation direction of the material of the disk (although, since the outward tips of almost all spiral arms point in the direction opposite of the disk rotation [47], the former can be used as a proxy measurement for the latter in many cases).

Figure 5.2: Galaxy classification options presented to human volunteers during the original Galaxy Zoo project (these buttons were arranged vertically on the Galaxy Zoo web site, archived at `http://zoo1.galaxyzoo.org`). We are primarily interested in the first two categories (clockwise spiral and anti-clockwise spiral) because they are used for galaxies with visible spiral arm structure. Note that galaxies classified as clockwise and anti-clockwise have Z-wise and S-wise patterns, respectively; we use the latter terminology in this work.

that spiral structure cannot be seen due to the viewing angle) or a non-spiral classification. For galaxies with higher discernibility rates, we expect our method to have better agreement with the Galaxy Zoo humans.

Table 5.1 gives winding direction agreement rates (within the 29,250-galaxy subset of Galaxy Zoo discussed at the beginning of this chapter) as a function of the minimum discernibility needed for a galaxy to be included in the comparison. Agreement rates are provided for several winding direction measures computed from our output. Table 5.1 also gives the percentage of galaxies included at each minimum discernibility, as well as the average discernibility within each of these subsets. In 19 out of the 29,250 images, our method did not report any arcs (arm segments), so these galaxies could not be included in the comparison. Although Galaxy Zoo more readily provides a measure of confidence in galaxy classifications (as quantified here by the discernibility), we can also include a rudimentary measure of certainty using our output: we can determine whether the two longest arcs have a length of at least 64 pixels (i.e., a quarter of the smaller image dimension) and agree in winding direction. Table 5.2 gives winding direction agreement rates with this criterion as an additional inclusion requirement.

As can be seen from Tables 5.1 and 5.2, even a simplistic winding direction measure, the majority vote, has reasonable agreement with Galaxy Zoo when computed from our output.

Min Discernibility	0	10	20	30	40	50	60	70	80	90	95	100
Inclusion Rate	99.9	99.8	99.7	99.4	98.9	98.0	95.7	91.0	79.4	52.5	32.1	12.5
Mean Discernibility	87.1	87.2	87.3	87.5	87.7	88.1	88.9	90.1	92.2	95.9	98.0	100.0
Majority Vote	80.6	80.6	80.6	80.7	80.7	80.8	81.0	81.4	82.2	83.8	85.3	85.5
Longest Arc	89.0	89.0	89.0	89.1	89.1	89.2	89.4	89.9	90.9	93.0	94.4	95.5
Length-weighted Vote	93.5	93.5	93.5	93.5	93.6	93.7	94.0	94.3	95.0	96.5	97.4	97.9

Table 5.1: Winding-direction agreement with human classifications from Galaxy Zoo 1. **Row 1:** the minimum proportion of human votes that the dominant winding direction must receive among the 6 categories from Galaxy Zoo 1. **Row 2:** the proportion of the 29,250 galaxies that met the above criterion and had arc detections from our method. **Row 3:** the mean discernibility rate among the galaxies included under these criteria. **Rows 4, 5, and 6:** agreement rates between Galaxy Zoo 1 and three methods of determining winding direction from our output.

Min Discernibility	0	10	20	30	40	50	60	70	80	90	95	100
Inclusion Rate	70.9	70.9	70.8	70.6	70.4	70.0	68.7	66.1	59.3	42.0	26.9	10.8
Mean Discernibility	88.8	88.9	89.0	89.1	89.3	89.5	90.2	91.1	92.9	96.1	98.1	100.0
Majority Vote	84.2	84.2	84.2	84.2	84.2	84.3	84.4	84.6	85.1	85.9	86.8	86.8
Longest Arc	97.9	98.0	98.0	98.0	98.0	98.1	98.2	98.3	98.6	99.0	99.3	99.6
Length-weighted Vote	97.7	97.7	97.7	97.7	97.8	97.8	97.9	98.1	98.3	98.8	99.1	99.4

Table 5.2: Winding-direction agreement with human classifications from Galaxy Zoo 1 using the same measures as Table 5.1, with the additional inclusion condition that the two longest arcs reported by our method must have a length of at least 64 pixels, and must agree in winding direction.

Adding arc-length information further improves accuracy, as seen with the agreement rates when considering only the longest arc, and when weighting by arc length. The arc-length-weighted vote is likely the best of these measures, agreeing with humans at least 93.5% of the time, and increasing to 97.9% with increasing human discernibility. When also requiring that the two longest arcs agree, agreement ranges from 97.7% to 99.4% for the arc-length-weighted vote. Under this additional criterion, agreement rates when using the longest arc alone are slightly higher than when using the arc-length-weighted vote, but the difference is small and the arc-agreement criterion preferentially selects galaxies where the longest arc is more reliable (since it has a minimum length and the second-longest arc also passes this minimum while agreeing in chirality).

The increase in agreement seen when using arc-length information suggests that our output contains small, spurious arcs (opposing winding directions may be present in the same galaxy, but this is likely very rare). The presence of these arcs is to some extent a necessary tradeoff due to our ability to represent arbitrary arm-segment structure: given what appears to be a small linear pattern in the image, we cannot dismiss it as noise simply because it does not fit a more restrictive model of spiral structure. Nevertheless, it may be possible to improve the noise-resistance of our method (beyond what we have already done) without sacrificing generality. When galaxy images have higher CCD noise, have low resolution, or are otherwise less well resolved, the risk of these noisy arcs increases. However, these problems can be partially mitigated by applying noise removal techniques such as the median filter discussed in Chapter 3. Furthermore, as seen with the discernibility variations in Tables 5.1 and 5.2, even human consistency varies with image quality (even within a sample of galaxies where humans reported that spiral structure was visible). If future sky surveys provide images with higher signal-to-noise ratios, we expect our output to improve accordingly. Additionally, as discussed in Section 3.7.2, the minimum cluster size is very small; this threshold was chosen because it provides more choice to the users of our method, not as a cutoff that maximizes agreement with Galaxy Zoo humans. Thus, improved agreement may be possible with more aggressive thresholding on arc length, cluster size, or other attributes.

The increase in agreement with humans when requiring agreement of two long arcs (i.e., the increase seen when comparing Tables 5.1 and 5.2) indicates that our output provides information about its expected accuracy for each galaxy. The arc-agreement criterion is not a perfect indicator; in pathological cases both long arcs may agree and wind in the incorrect direction (see Figure 5.3 and the related discussion), and one-arm and flocculent (highly fragmented) spiral patterns often cannot satisfy this long-arc-agreement criterion even if these patterns are described perfectly. Using other information from our output (e.g., taking all arcs into account and considering other properties of the arcs, such as their radial extent), other measures of confidence may be possible.

With high human discernibility (where we expect the human classifications to be highly reliable), our agreement rate is very high (97.9% for 100% human discernibility). However, even with 100% human discernibility and our requirement for two fairly-long agreeing arcs, there are a few cases (20 out of 3145, i.e., about 0.6%) where we still do not agree with human classifications. It is useful to study these cases, shown in Figure 5.3, because they illustrate ways in which our method can occasionally fail. For some galaxies, the image standardization step (Section 3.3) "zooms in" too far because the ellipse fit was too tight. This can occur for extremely faint spiral arms, and when it does happen, most or all of the spiral arm region is not visible in the standardized image, losing much or all of the information needed to determine spiral structure. The effect has been substantially reduced with an appropriate brightness transformation (in particular, by selecting the linear-to-logarithmic transition parameter β, discussed in Section 3.2, via a percentile level that is as low as possible while still retaining meaningful brightness variations, so that faint details of the galaxy are brought out as much as possible), but improving the brightness transformation did not entirely eliminate this effect.

Other winding direction disagreements occur due to arcs that are either spurious, or are not strongly informative about winding direction. In some of these cases, the two longest arcs have the "correct" winding direction but are outweighed by other arcs, while in other cases, the two longest arcs are spurious (at least in terms of winding direction) and outweigh the correct-direction arcs. Some of the incorrect-direction arcs are reasonable in terms of overall galaxy structure, but do not capture winding direction information. These arcs often describe low-pitch-angle, ring-like structures (and thus structures without a well-defined winding direction) at the galaxy edge or near the bulge (see, for example, the third galaxy from the left in the third row of Figure 5.3), but can include other phenomena such as opposite-chirality arms (or arm-like filaments), or severe bends in the spiral arms. In some other cases, the arcs are more likely to be due to image noise; this factor was discussed earlier in this section. Figure 5.3 includes some especially severe cases of input image noise, such

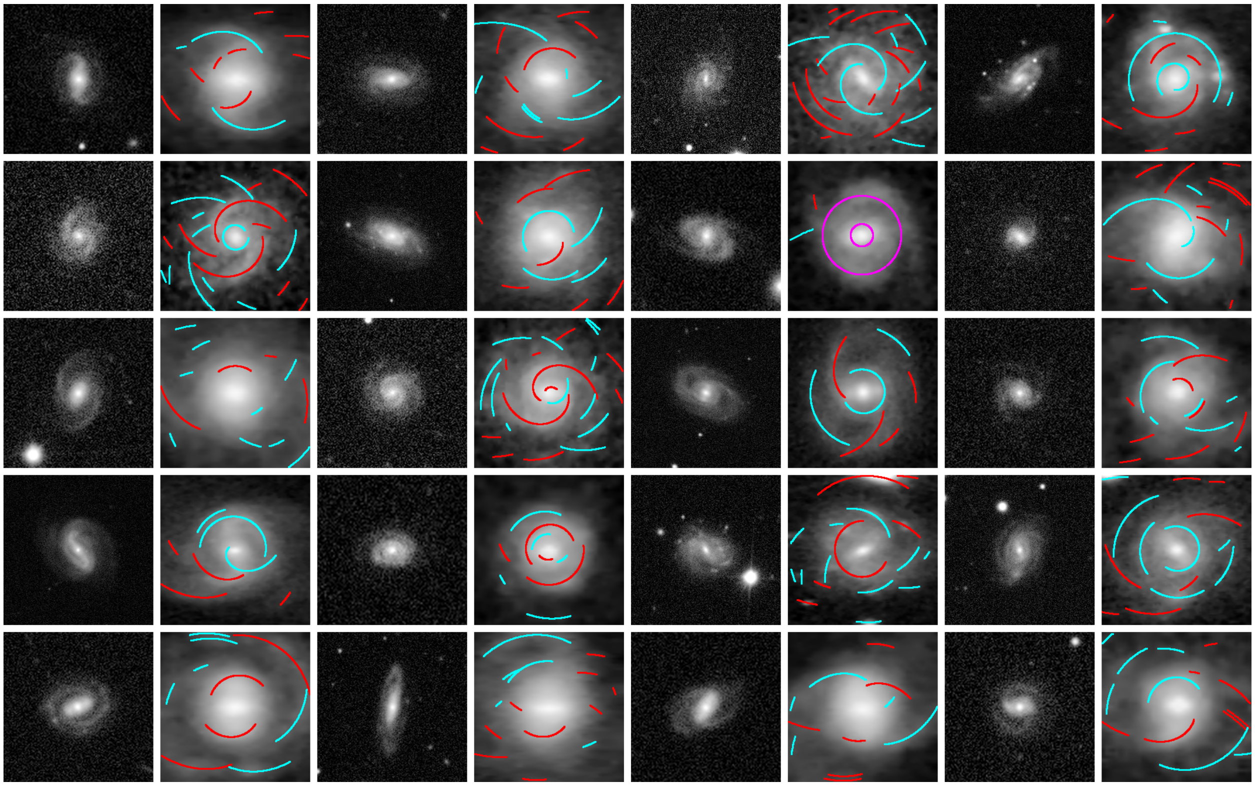


Figure 5.3: Input images and final detected arcs (overlaid on the standardized images) for the 20 galaxies (i.e, 0.6% of the galaxies) with 100% human discernability (and thus 100% human agreement) and where our method produced two fairly-long arcs that agreed in winding direction, but where our winding direction did not agree with the human classification. Our winding direction measurements agree with the human classifications for the other 3125 galaxies meeting these arc-agreement and human discernibility criteria. For visibility of the input galaxies, the input images are cropped at 25% on each side when displayed here.

as the third galaxy in the top row and the first galaxy in the second row (from the top).

In one case shown in this Figure, a galaxy could be interpreted either as a two-arm galaxy or as a ring galaxy. The code originally recognized the two separate arms (before the secondary merging step discussed in Section 3.7.2), but merged them because the ring fit was good enough (if two clusters can be reasonably explained with one arc, our code currently merges them). This ring does not have a winding direction, and absorbs all of the "good" winding direction votes. In response, one could try reducing the threshold (increasing the stringency) of the arc fit merge ratio (Equation 3.19) for all arcs or for merges into rings, or prohibiting merges into rings if one or both arcs have a pitch angle higher than a threshold. However, like many tempting tweaks to the method described in Chapter 3, such a change risks causing more problems than it solves, especially since these failure cases are rare and unusual.

Returning to our overall assessment of winding direction agreement, it is also interesting to consider how well our winding direction reliability compares to the average human volunteer in Galaxy Zoo. For this we can consider average agreement (the proportion of galaxies where the winding direction given by our code agrees with the majority human vote) and average discernibility (the average proportion of human votes for the dominant winding direction category). Since the galaxies sometimes may not be sufficiently well resolved to have a clearly correct answer for winding direction, and since humans had the opportunity to indicate uncertainty by choosing a category other than clockwise or anti-clockwise (Z-wise or S-wise) spiral (with no indication of whether they would have agreed with the majority vote), we cannot exactly compare winding-direction accuracy rates. However, average discernibility still provides a measure of how well each human perceives winding direction; a vote for the "wrong" (non-majority) winding direction and a vote for a non-winding-direction category are both evidence that the galaxy's winding direction is difficult for humans to determine. Additionally, although discernibility is computed on a per-galaxy basis and our agreement with Galaxy Zoo is computed across the set of galaxies, the two measures are still directly comparable in this regard. Within a set of galaxies (for example, the subset corresponding to any of the numerical columns of Table 5.1 or 5.2), the mean discernibility gives the probability that a randomly selected human classifier will agree with the majority vote for a randomly selected galaxy, while the agreement rate gives the probability that our winding direction classification for a randomly selected galaxy will agree with the majority human winding-direction vote. For a more detailed argument as to the comparability of the mean discernibility of galaxies and mean agreement with humans despite discernibility being computed for each galaxy and agreement being computed across galaxies, see Davis and Hayes (2014) [13].

Comparing the bottom row of Table 5.1 with the "Mean Discernibility" row of the same table, and assuming the majority human vote within the two winding direction categories gives the "correct" answer, we find that our length-weighted average vote agrees with the human majority vote for winding direction more often than the average human voter does.[2] The two exceptions are the subsets of galaxies with the highest discernibility cutoffs (95% and 100%) — sets that, by definition, favor the galaxies with the highest human reliability. In these sets, our method agrees with the human majority vote about 98% of the time. Further investigating the observation that human discernbility is only higher than our agreement (with the majority human vote) for the highest discernibility cutoffs, we find that average discernability equals average agreement at a point between the 90% and 95% discernibility cutoffs. Figure 5.4 illustrates this crossover point. When the minimum discernibility is 92.9%, our method agrees with the majority winding-direction vote at that same rate that humans perceive the "correct" winding direction; this agreement rate is about 97.1%. At this point, the image inclusion rate is 41.78%, so for most images in our comparison set, our agreement with the majority winding direction vote exceeds the rate at which humans reliably perceive the winding direction.

Longo [42] also provided winding direction measurements for a set of galaxies, where each galaxy was viewed only once by one of five student volunteers. Many of these galaxies (about 52%) were also used in our comparisons with Galaxy Zoo, so for the galaxies in this intersection we can compare our winding direction with those reported in Longo [42]. In Longo's survey, each galaxy could be classified as Left (i.e., a "left-handed" winding direction), Right, or Uncertain. The volunteers were instructed to use the latter category unless winding direction was clear, leaving about 15% of the original sample in [42]. Thus, this set is similar to a higher-discernibility subset of the Galaxy Zoo sample. As shown in Table 5.3, our agreement rate is also similar; the arc-length-weighted vote achieves 95.4%

---

[2] We repeat our earlier caveat: although a human vote for a non-winding-direction category indicates difficulty perceiving a winding direction, human discernibility and our agreement are not exactly comparable because our method is forced to choose a winding direction.

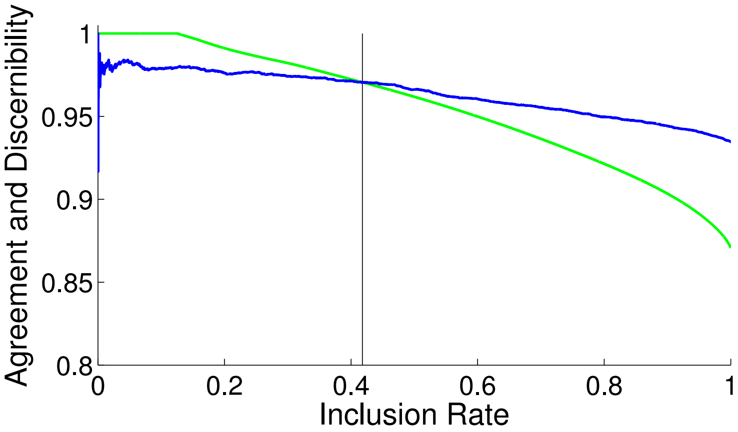


Figure 5.4: As we include progressively more galaxies from our Galaxy Zoo comparison sample in order of human discernibility, average human discernibility in winding direction (green) initially exceeds our average agreement with the majority winding-direction vote (blue), but average human discernibility decreases faster than our average agreement with the majority human vote. The crossover point occurs at a 41.78% inclusion rate, so in most of the sample, our agreement with the majority human vote exceeds the rate at which humans reliably perceive the majority-vote winding direction.

	All	**Longest Agree**
Majority Vote	83.0	85.6
Longest Arc	91.7	98.5
Length-weighted Vote	95.4	98.3

Table 5.3: Agreement between our winding direction and that of Longo [42] for the 13,331 galaxies intersecting the set of 29,250 galaxies used for comparisons with Galaxy Zoo. The last column uses the same arc-agreement criterion as Table 5.2.

agreement across cases where the human classifier decided that the winding direction was clear. When requiring that our output had at least two fairly-long arcs with the two longest agreeing in winding direction (which occurred about 76.3% of the time), agreement increases to 98.3%, which matches what we found with the Galaxy Zoo classifications for a comparable discernibility rate. Since each galaxy was only seen by one observer in Longo (2011) [42], we cannot measure human agreement in Longo's results.

5.2 Spiral Arm Tightness (Pitch Angle)

When spiral arm tightness is measured, it is generally given as the pitch angle, i.e., the angle a spiral arm makes with a circle centered at the galaxy center. For each detected arc in a spiral galaxy, our method readily provides a pitch angle: it is the arctangent of the a parameter of the logarithmic spiral arc (Equation 3.1) fit to the arm segment. If two or more spiral arms within the same galaxy have different pitch angles, the pitch angle of the galaxy as a whole is not uniquely defined. In the pitch angle comparisons throughout this section, we calculate the pitch angle of a galaxy as the arc-length-weighted mean of all arcs that agree with the dominant winding direction. We use this measure for several reasons. First, we have found arc length to be evidence of the arc's importance to overall galaxy structure (for example, consider the winding direction agreement rates in Table 5.1 for the arc-length-weighted vote as compared to a simple majority vote). Longer arcs are also less likely to be due to noise (but we cannot eliminate all short arcs because flocculent galaxies naturally have smaller, fragmentary arms). We only use arcs agreeing with the dominant winding direction because arcs with the opposite winding direction are more likely to be noise-induced, and even in cases where opposite-direction arm segments exist, their tightness may not be directly comparable with arms of the dominant winding direction.

We first compare our pitch angles with arm tightness classifications given in Galaxy Zoo 2 [67]. Here, where human classifiers indicated that there was "any sign of a spiral arm pattern," they were asked whether the arms were tight, medium, or loose (see Figure 5.5). Although examples and illustrations were given for each category, choices are not completely precise, and the categories are too coarse for a direct correspondence with our measurements. However, we can examine the distribution as a whole.

Figure 5.6 shows the relationship between our measured pitch angle and the proportion of galaxies receiving a majority weighted human vote for Tight, Medium, or Loose (the vote


How tightly wound do the spiral arms appear?

Figure 5.5: Reference icons presented to human classifiers when asked about spiral arm tightness during the Galaxy Zoo 2 project, as seen on `http://zoo2.galaxyzoo.org`.

weights are determined by consistency with other humans; see [67]). As can be seen from the dashed lines, galaxies where we measure a low pitch angle usually have majority human votes for Tight, while most of the remaining galaxies in this range had majority votes for Medium. 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.

In the top human agreement quartile (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 5.6. Also, since majority votes for Medium were far less common with increased human agreement (majority votes for Medium decreased from 49.3% to 17.2% when using only the highest human agreement quartile, and the proportions of Medium-majority votes are also much lower in Figure 5.6), it seems likely that this choice was frequently used to indicate uncertainty, perhaps due to low galaxy resolution and galaxies with arms of varying tightness. Consequently, it is reasonable that majority-Medium galaxies spread across a wide range of our measured pitch angles. Even if Medium-majority galaxies are disfavored by the entropy measure by 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 classifications stems from human uncertainty. In all, then, we see a clear association between Galaxy Zoo 2 tightness classifications and our measurements, with this association strengthening as human agreement increases.

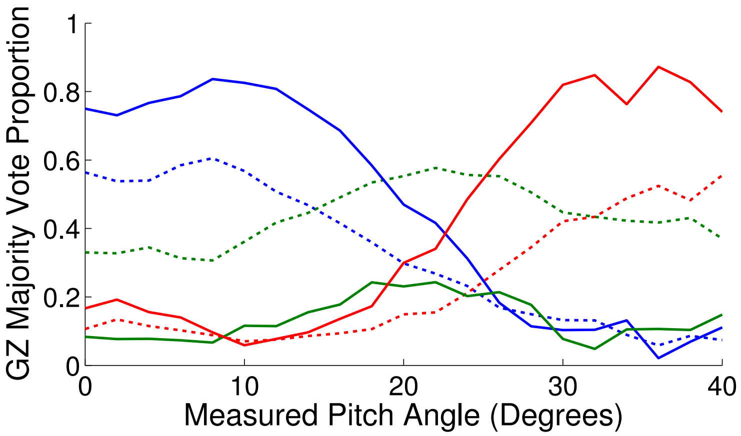


Figure 5.6: Proportion of galaxies receiving a majority vote for Tight (*blue*), Medium (*green*) or Loose (*red*) as a function of our measured pitch angle, over all images tested from Galaxy Zoo (*dashed*) and for the top human agreement quartile (*solid*). Pitch angles are binned with width 2 degrees from 0 to 40, with one more bin for pitch angles above 40 degrees (due to low sample size). Overall, 35.7% of galaxies had a majority vote for Tight, 49.3% had a majority vote for Medium, and 15.1% had a majority vote for Loose. In the top human agreement quartile, these rates were 55.8%, 17.2%, and 27.0% for Tight, Medium, and Loose, respectively.

Figure 5.7 gives the number of galaxies in each pitch angle bin for both the full set of 29,250 galaxies and for the top human agreement quartile (corresponding to the dashed and solid lines of Figure 5.6, respectively). Comparing these distributions with Figure 5.6, we note that the majority human votes for Tight and Loose have proportions closer to equal where we measure "typical" pitch angles. The peak in our pitch angle frequencies occurs at a lower pitch angle than the Tight/Loose crossover points in Figure 5.6, but this is likely because there are fewer galaxies with a majority vote for Loose than for Tight, so a higher pitch angle is required before Loose-majority-vote galaxies overtake Tight-majority-vote galaxies. From visual inspection we can also observe that the pitch angle distribution of galaxies in the top human agreement quartile is similar to the distribution of all galaxies used for comparison.

We also compare our measured pitch angles with values calculated using Fourier analysis by Davis et al. [11] (no relation). Here, we perform the comparison using images from the Carnegie-Irvine Galaxy Survey [25]; this allows inclusion of most of the galaxies measured by Davis et al. [11]. We directly use PNG images where foreground-star-containing regions have

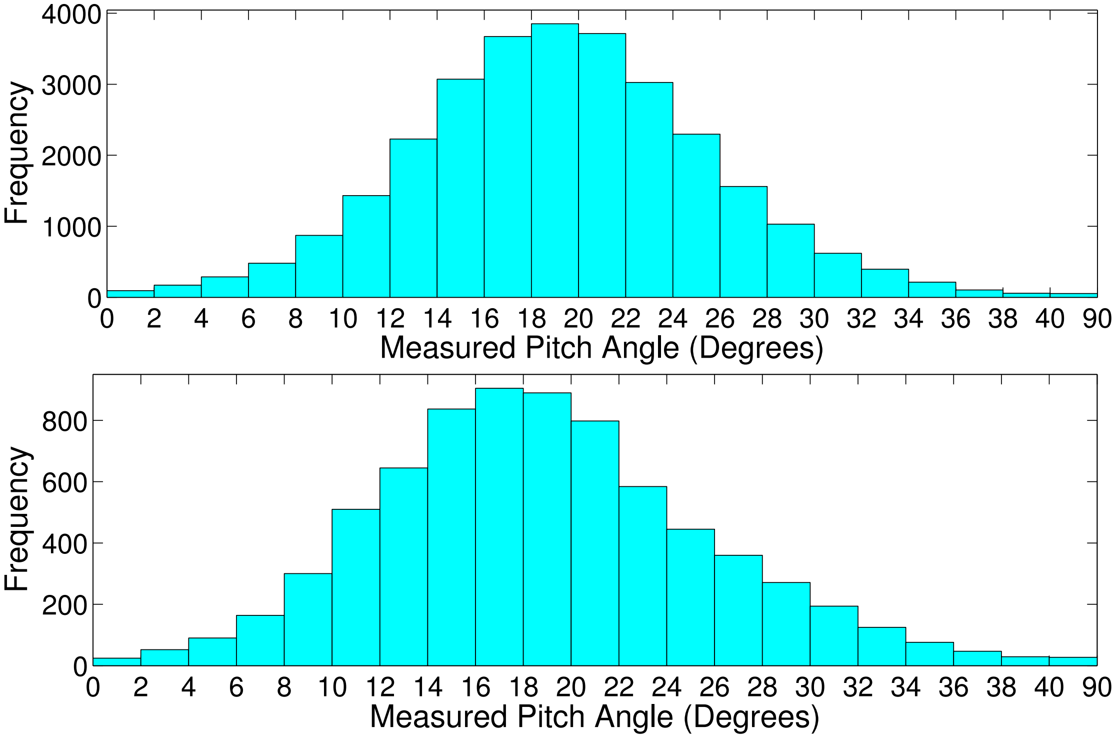


Figure 5.7: The distribution of our measured pitch angles across all 29,250 galaxies used in our comparison with Galaxy Zoo (*top*) and for the galaxies in the top human agreement quartile (*bottom*). Note that all pitch angles above 40 are combined into the same bin.

been patched with substitute brightness values. We could instead use FITS images with our own PNG conversion, with similar (but not identical) results, but we experienced difficulty getting SExtractor output usable for the star removal discussed in Section 3.4 (getting such output is likely possible by adjusting SExtractor parameters from their defaults in order to better match the properties of this data set), and we also take the opportunity to demonstrate the use of our method on non-FITS inputs. The left panel of Figure 5.8 compares our pitch angles with those measured in [11] for the I (infrared) band. We see a reasonable agreement in most cases (with some scatter, discussed below). The galaxy farthest from equality in measured pitch angle, NGC 5054, has an extreme pitch angle measurement in Davis et al [11]. We also measure a large pitch angle, but not to the same extent, and our agreement with their B (blue) band pitch angle is much closer. This galaxy has a large pitch angle difference among its arms, and is revisited later in this section and in Figure 5.10.

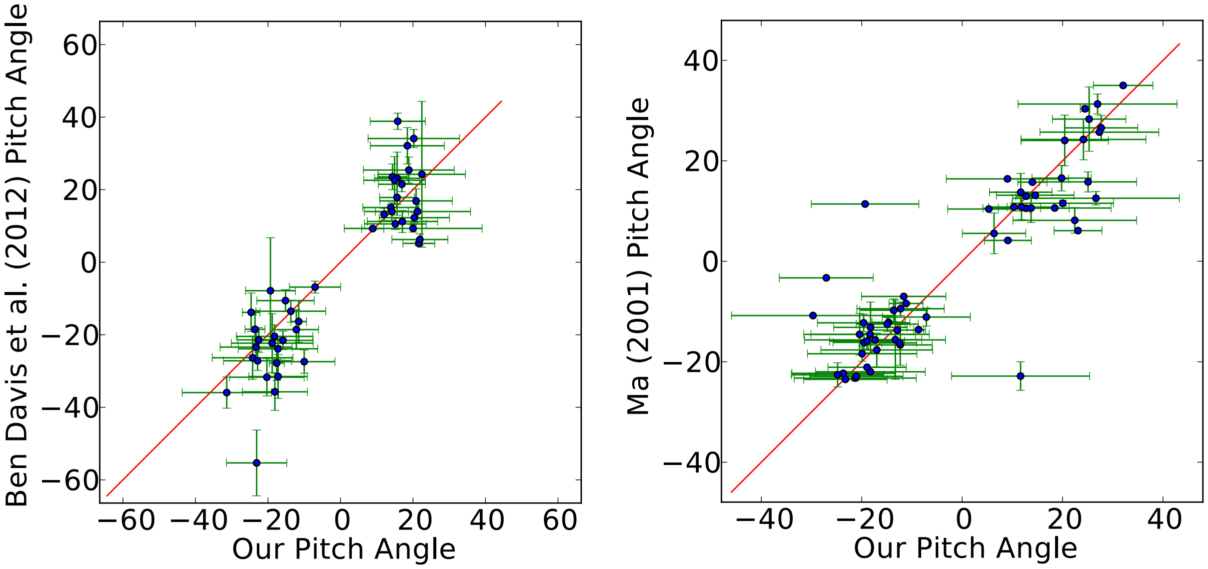


Figure 5.8: Comparison of our measured pitch angles with those from Davis et al. (no relation) [11] for the I band, left, and with those reported by Ma [44], right. The red lines plot $y = x$. In both cases, the vertical axis is the pitch angle measured by the other authors, and the horizontal axis is our measured pitch angle. Our error bars are the arc-length-weighted standard deviation of all arcs agreeing with the dominant winding direction (as determined by arc-length-weighted vote). Davis et al. describe their own method for determining error values [11] which we include here; for the error from Ma [44] we take the difference between the two measured arcs (when two measurements are available).

We perform a similar comparison with Ma [44], where pitch angles were determined by manually selecting points along spiral arms, and then fitting one or two logarithmic spiral arcs (depending on visual resolvability). For this comparison we ran our code on images from the POSS II survey [53], since this was the image set used by Ma [44]. This survey used photographic plates rather than CCD imaging. Like the Galaxy Zoo comparison set, we applied a brightness transformation with the parameters discussed in Section 3.2. Since the use of photographic plates had an effect similar to a logarithmic transformation of brightness values, our brightness transformation did not have (and did not need to have) as strong of an effect on extreme brightness values. However, the distribution of brightness values was more concentrated toward the high end, so applying the unsharp mask (Section 3.5.2) created more brightness clipping than normal. Thus, to cope with the use of photographic plates,

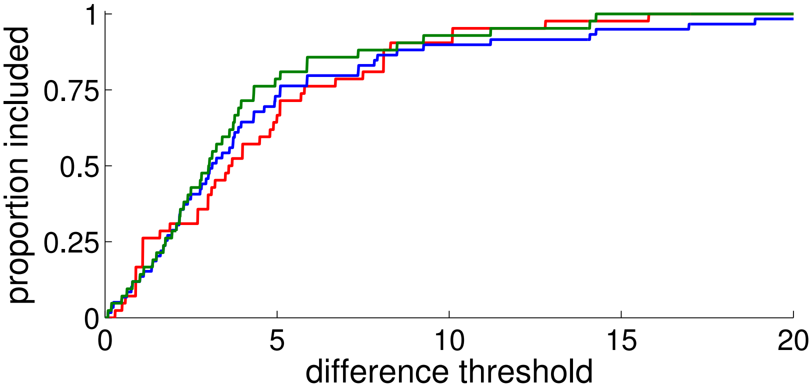


Figure 5.9: Cumulative distribution of within-galaxy pitch angle discrepancies (ignoring sign) from Ma [44] (in red) for galaxies where two arcs are measured, and between our measurements and the measurements in [44] (blue for the full comparison set, and green for the subset where two arc measurements are available from [44]). Larger values mean stronger agreement; if all measured values were the same, the inclusion rate would have a constant value of 1.0.

we reduce the unsharp mask amount parameter[3] (Section 3.5.2) from 6 to 3. Using the original unsharp mask amount degrades the results slightly, but not severely. The right panel of Figure 5.8 compares our pitch angles with those in Ma [44]; like our comparison with Davis et al. [11], there is a rough correspondence, but with some scatter. There are two galaxies, NGC271 and NGC6956, where we disagree in winding direction; in these two cases, our winding direction is incorrect. NGC271 has ring-like structure on each side of the ends of its bar, producing plausible arcs of similar pitch angle in both winding directions; the NGC6956 image is contaminated with stars, including several stars visually arranged near one of the spiral arms.

Both quantitative pitch angle comparisons have some scatter. Although this scatter may seem to be a concern, it is possible for the pitch angle to vary significantly between arms in the same spiral galaxy. Since Ma [44] measured only one or two arms in each galaxy, and since Davis et al. [11] measured only one dominant pitch angle, perhaps much of the scatter could be explained by inter-arm differences in each galaxy, rather than by differences as a

---

[3] This manual tuning is at the level of the entire image set; per-image supervision is still entirely unnecessary.

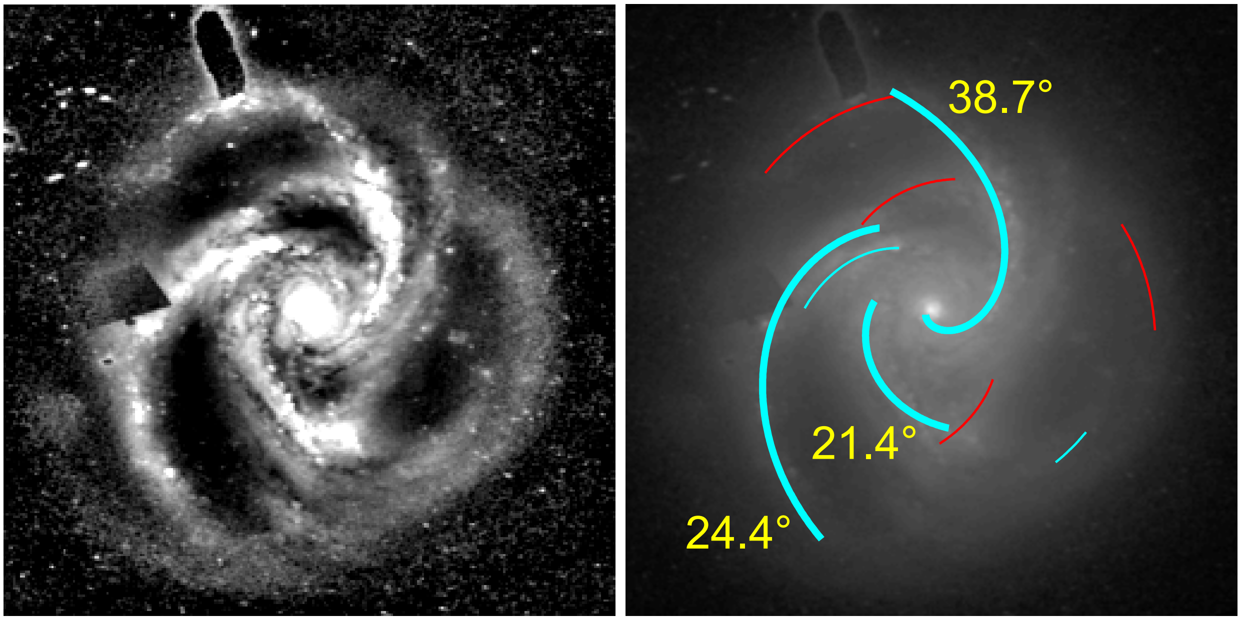


Figure 5.10: From our output we find that pitch angles can vary substantially within the same galaxy, as shown here for NGC 5054. The left panel shows the image, obtained as a PNG input from the Carnegie-Irvine Nearby Galaxy Survey [25] with star regions replaced with substitute brightness values, after image standardization (Section 3.3) and unsharp masking (Section 3.5.2). The right panel shows the arcs detected by our method, along with pitch angles for the three longest arcs that agree with the dominant winding direction. Compared to the other two arcs, the upper arc differs in pitch angle by about 15 degrees. Even without the innermost region of the upper arc (a discontinuous region joined during the secondary merging step described in Section 3.7.2), the pitch angle of the upper arc remains similar — about 37.8 degrees.

function of method. To test this hypothesis, Figure 5.9 plots the cumulative distribution of pitch angle discrepancies between the two arcs in one galaxy (when available) from Ma [44], vs. discrepancies between Ma's method and our method. All three curves are similar, with between-method agreement often being higher than within-galaxy agreement. It is thus likely that much (but not all) of the scatter in Figure 5.8 arises from within-galaxy arm variation rather than between-method measurement variation. Figure 5.10 illustrates one case of within-galaxy pitch angle variation. The three main arms are fit well, but one of the arms differs in fitted pitch angle by about 15 degrees. We see pitch angle differences in other galaxies as well (though usually not as severely as in this case).

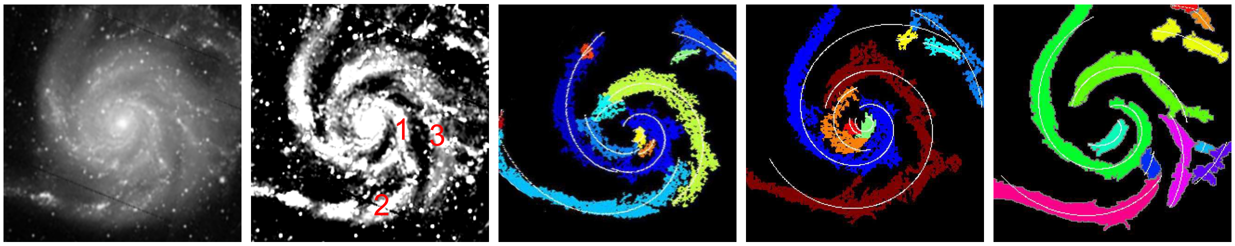
5.3 Spiral Arm Counts

As mentioned in Section 3.1, the concept of a "spiral arm" can be ambiguous. Consequently, the notion of an arm count is often not well defined. Human viewers (including Galaxy Zoo classifiers) may implicitly or explicitly have different criteria for what properties constitute a spiral arm (e.g., shape, coherence, length, brightness), and how prominent a spiral arm must be in terms of these properties. Differences in these criteria affect the number of arms perceived in the image. At suspected junctions of spiral arms (for example, forks, bends, and dust lanes), it can also be unclear which spiral arm, if any, is a continuation of another. This situation, illustrated in Figure 5.11, can also affect the arm count. The instructions given to Galaxy Zoo human classifiers[4] also note the ambiguity in spiral arm count, stating "How many distinct arms can you see? The correct answer is often not very clear — spiral arms can be messy, so a little imagination helps. If you're uncertain, say between 2 or 3, just give your best guess. If you can see there are spiral arms, but you really can't tell how many, then you have the option of saying so."

Due to the ambiguities in counting spiral arms, the task of reproducing human classifications is related to, but distinct from, the task of tracing spiral arm structure. The former does not require as much output detail, and can exploit correlations between image features (and/or other signals) and the category labels, without ever needing to directly model spiral structure. In contrast, our output attempts to capture the details and subtleties of spiral arm structure, including ambiguity in the spiral arms; this is why we made the distinction between "spiral arms" and "spiral arm segments" in Section 3.1. From our output the user can then define and calculate structure measures appropriate to the task at hand. In the case of arm count, if a researcher is only interested in spiral arms that start near the core and end near the disk, they can apply this criterion, while a broader interpretation of spiral arms would be

---

[4] See the section "Q: How many spiral arms are there?" from `http://zoo2.galaxyzoo.org/how_to_take_part`



(a) (b) (c) (d) (e)

Figure 5.11: Three different interpretations of the spiral structure in M101. All comprise reasonable interpretations of the structure. Part (a) gives the standardized image (as described in Section 3.3). Part (b) displays the image after unsharp masking (Section 3.5.2); 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 explicitly refers to forks. The colored images are from different stages in the development of our method. Part (c) gives output from an old version of the code, where the three segments happen to be separated (blue, cyan, and olive pixel clusters). Part (d) gives output from an intermediate (but still quite old) version of the code, where segment 2 has been interpreted as a continuation of segment 3, jumping over the gap between them. The single logarithmic spiral arc spanning the two arguably fits reasonably well, suggesting perhaps that 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. The arc at the top of part (d) does not fit well due to an old bug (resolved long ago) in setting up the least squares fitting for multi-revolution arcs; the "real" logarithmic spiral arc fit matches much better. Part (e) gives output from the most recent version of the code, in which the logarithmic spiral arcs more stringently fit (in the least-squares sense) each cluster of pixels. Segment 3 has been split into two clusters, suggesting that a fork exists at the split point.

more appropriate in other cases and can also be applied to our output. For the purpose of reproducing human classifications of spiral arm count (e.g., from Galaxy Zoo 2), we expect our output to produce useful information, but it is not immediately clear which mapping from our output to human classifications would be best. Furthermore, given a fixed mapping, it is not clear whether (or to what extent) arm count differences are due to limitations in our method (or human classifications), or due to differences in how an arm was defined by humans and by the mapping. Since we are primarily interested in quantitatively describing spiral structure, and since we have already compared with other measures (especially winding direction; see Section 5.1) that have a well-defined answer much more frequently, we leave for future work the task of more precisely reproducing human arm-count classifications from

Figure 5.12: Reference icons presented to human classifiers when asked about spiral arm count during the Galaxy Zoo project, as seen on `http://zoo2.galaxyzoo.org`.

our output. However, despite the ambiguities of arm count definition (and thus in matches and mismatches of arm count), we can still compute several simplistic measures of arm count from our output and then examine the correspondence between these measures and human classifications. It is likely that other measures (based on our output) could have improved agreement rates with Galaxy Zoo 2 classifications, but we still expect these simple measures to have some correspondence, especially for cases with higher human agreement.

During Galaxy Zoo 2, 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 1, 2, 3, 4, or more than 4 arms, along with a “Can’t tell” option. These options are displayed in Figure 5.12. For each galaxy, we consider the arm count determined from Galaxy Zoo to be the category with the highest fraction of weighted votes (the vote weights are determined by consistency with other humans, as was the case with arm tightness, and as described in [67]). 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 arms with length (in pixels) exceeding a threshold.

For the same reasons discussed in Section 5.2, we only include arcs that agree with the galaxy's dominant winding direction, as determined by the arc-length-weighted vote. For correspondence with Galaxy Zoo 2 categorizations, we group all counts above 4 within a single "more than 4 arms" category. Our (simplistic) arm count calculation method can place galaxies in a "0 arms" (really "all short arms") category, which has no equivalent in the Galaxy Zoo 2 classifications. Likewise, the Galaxy Zoo 2 "Can't tell" category has no equivalent in the categories computed from our output.

Among galaxies, the category for "2 arms" is much more popular (in terms of the Galaxy Zoo 2 maximum vote) than the other categories, so we must make sure that indicators of performance on other types of spiral galaxies are not overwhelmed by the relatively large number of 2-armed spirals.[5] To do so, we consider each Galaxy Zoo 2 arm count category separately. We partition the Galaxy Zoo comparison set (the same one used in Sections 5.1 and 5.2 when comparing with Galaxy Zoo) according to the Galaxy Zoo 2 arm count category receiving the maximum human vote for each galaxy. For each such group of galaxies (each corresponding to a Galaxy Zoo 2 arm count category), we find the proportion of galaxies our method assigns to each arm count category. These distributions are given in Figure 5.13. The rows correspond (from top to bottom) to arc length thresholds of 65 and 90. The left column includes all galaxies in the Galaxy Zoo comparison set. The right column uses only the galaxies in the highest quartile of human agreement (as measured by the Shannon entropy across human arm-count votes).

Regardless of threshold on arc length (i.e., the minimum length needed to be counted as a "spiral arm") and whether or not we include all galaxies or just the top human-agreement quartile, our method tends to give the lowest arm counts to galaxies classified as "1 arm" by Galaxy Zoo 2 (the distribution of these galaxies is given in red), and for each subsequent

---

[5] Incorporating a bias toward 2-arm galaxies would be appropriate when applying supervised machine learning methods to reproduce Galaxy Zoo 2 classifications from our output, but to preserve generality in our structure descriptions we do not wish to incorporate such a bias on the galaxy level, and for the comparison here we are interested in behavior for all arm counts.

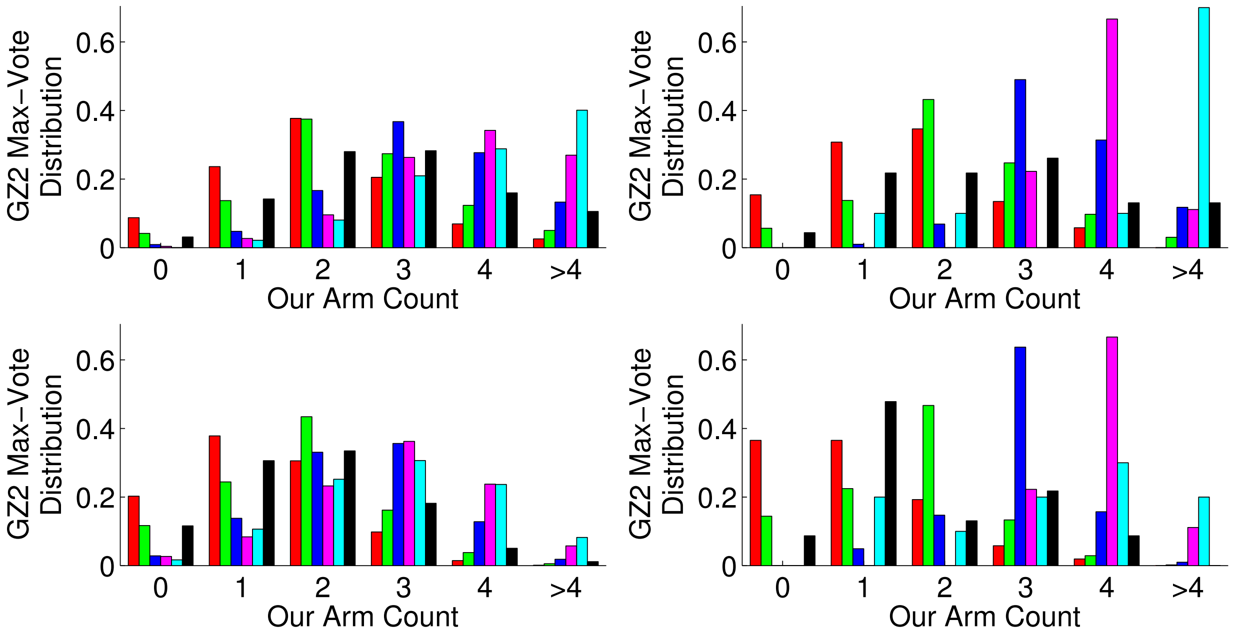


Figure 5.13: Distributions across arm count categorizations (0, 1, 2, 3, 4, and more than 4 arms) determined by our method for galaxies with a maximum human vote for “1 arm” (red), “2 arms” (green), “3 arms” (blue), “4 arms” (magenta), “More than 4 arms” (cyan), and “Can’t tell” (black). We use the maximum human vote because some galaxies do not have a majority human vote for any one category. In all of the arm counts produced from our output for comparison here, we only consider arcs that agree with the dominant winding direction, as determined by an arc-length-weighted vote. We then determine arm count by counting arcs with length of at least 65 pixels (top) or 90 pixels (bottom). The left column displays the distributions when all galaxies within our Galaxy Zoo comparison set are included. The right column includes only galaxies within the top arm-count agreement quartile (lowest Shannon entropy quartile).

Galaxy Zoo 2 arm-count category (ordered by number of arms and excluding “Can’t tell” because it does not indicate an arm count), our method tends to give higher arm counts (i.e., the distributions tend further to the right).

When including all galaxies in the Galaxy Zoo comparison sample, 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 peaks at N arms). Smaller arc-length thresholds (e.g., the top row of Figure 5.13) produce this matching for higher arm counts, and larger arc-length thresholds (e.g., the bottom row of Figure 5.13) produce this matching for lower arm counts. Thus, lower arc-length thresholds often over-count arms (relative to human counts) in galaxies were humans report few arms, and higher arc-length thresholds often under-count arms in galaxies where humans report more arms. These mismatches are likely due to a combination of arm-detection errors or other inherent differences between our method and human perception, and subtleties in human classifications not captured by a simple arc-length threshold.

When small regions appear linear due to noise or other imaging artifacts, we may over-count arms (especially at low arc-length thresholds). If the arm (or part of it) is too dim or otherwise too poorly resolved to be picked up by the orientation filters (and thus the pixel clustering), or if a dust lane obscures part of a spiral arm, then arms can be under-counted (if the detected region of the arm is too short) or over-counted (if the non-detected part of the arm causes a split into two segments that each pass the length threshold). Under- and over-counts (relative to Galaxy Zoo 2 classifications) due to dim or obscured arms can be valid interpretations in some cases, depending on how one defines a spiral arm.

Subtleties in human classification are more difficult to examine, but we speculate that some of the mismatches (of distribution peaks and on a per-galaxy basis) may be due to an overshadowing effect in human classifications: some arms may be implicitly considered dominant by being longer than other arms in the galaxy, and non-dominant if they are shorter than other arms or are attached to another, longer arm (e.g., as a fork or bend). If human classifiers consider the dominant arms to be the "real" arms, they may not count the non-dominant ones, even if an arm of similar length would be counted in a galaxy without a dominance pattern. If galaxies with more arms (of similar length) tend to have shorter arms and are less likely to have a dominance pattern,[6] then a shorter arc-length threshold will correctly detect more of the shorter arms in many-armed galaxies, but will tend to over-count (with respect to human classifications) arms in galaxies where humans perceive fewer arms, due to differences in whether the minor (non-dominant) arms are counted. A larger arc-length threshold would be more likely to "correctly" count only the dominant arms, but miss more

---

[6] This is plausible in part because "grand design" spirals tend to have two long coherent arms, while flocculent spirals tend to have a larger number of shorter arms.

of the shorter arms characteristic of many-armed spiral galaxies. Additionally, human classifiers may consider other factors that are not captured by an arc-length threshold (but may be available from our method already), such as the relative brightness of an arm.

Other reasons for arm-count mismatches may be a combination of human arm count decision subtleties and the way our output represents spiral arms. For example, if an arm has a sharp bend, our method may split it into two arm segments while humans may consider the two sides of the bend to be the same arm. In many astronomical applications, it may be useful to represent the arm as two separate segments with information (e.g., adjacency at the ends) available from our output that could signal membership in the same arm under some definitions. In cases like these, where increased optimality for one purpose (e.g., Galaxy Zoo 2 arm count agreement) could reduce available information, suitability, or flexibility for other astronomical applications, we have chosen to keep our output suitable for a broader array of purposes, and allow others to calculate application-oriented measures from our output. As discussed at the beginning of this section, we are interested in arm count agreement to the extent that it measures the suitability of our output for general astronomical purposes, and consider application-specific optimizations using our output to be beyond the scope of the current work.

Although it is currently difficult to measure the proportion of arm-count mismatches attributable to errors in our method versus other factors such as subtleties in human classifications not captured by an arm count threshold, we can consider the changes we see when only considering the galaxies in the top quartile of human agreement (as measured by Shannon entropy). As with the left column, the right column of Figure 5.13 shows distributions across our arm-count values for galaxies partitioned according to which arm-count category received the maximum Galaxy Zoo 2 vote, except that only the galaxies in the top human agreement quartile are included. With this inclusion criterion, all distributions except for "1 arm" peak at the matching arm count (for example, of the galaxies with a maximum hu-

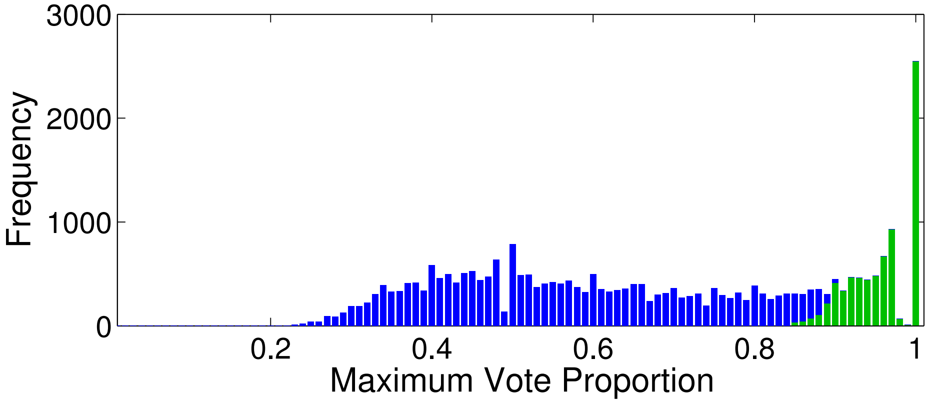


Figure 5.14: The distribution (across our Galaxy Zoo comparison set) of the weighted vote that each galaxy received for the arm count category with the maximum human vote. The instances within the top human agreement quartile (as measured by Shannon entropy) are in green.

man vote for 4 arms, our method counts 4 arms in most cases, as shown in the distribution displayed in magenta), regardless of the arc-length threshold used. The distributions are also more concentrated around their peaks. There are still arm-count disagreements at the galaxy level, and these can still be due to either errors in our method or nuances in human classification not captured by the arc length threshold (but potentially captured by other information included in our output), but it is interesting to note that our agreement with humans increases substantially when the humans agree with each other.

Figure 5.14 assesses the extent to which humans agree with each other. We can 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. Table 5.4 gives the percentage of galaxies assigned to each arm count category by human classifiers and by our three arc-length thresholds, for the full Galaxy Zoo comparison set and for the top human agreement quartile. Both the human classifications and our methods assign the most galaxies to the "2 arms" category, and both assign progressively fewer galaxies to categories with progressively more or fewer arms. However, human classifications are more

Measurement	Top?	0 Arms	1 Arm	2 Arms	3 Arms	4 Arms	>4 Arms	? Arms
DWD, $\geq$065px	**N**	3.56	12.67	32.33	28.06	15.25	8.13	0.00
	Y	5.62	13.71	42.45	24.91	10.06	3.24	0.00
DWD, $\geq$090px	**N**	10.61	24.20	39.27	19.06	5.78	1.08	0.00
	Y	14.32	22.37	45.86	14.03	3.19	0.25	0.00
GZ2 Max Vote	**N**	0.00	2.82	64.04	9.42	2.69	2.25	18.77
	Y	0.00	0.71	97.32	1.39	0.12	0.14	0.31

Table 5.4: Percentages of galaxies assigned to each arm count category as a function of measurement method (first column) and whether we only include the top human agreement quartile (second column). In the arm counts produced from our output, we only include arcs that agree with the dominant winding direction (DWD), and also require that the arcs have a minimum length in pixels. Note that Galaxy Zoo 2 did not have an option to vote for "0 arms" (i.e., "only short arms"), and our method does not produce classifications in the "Can't tell" ("? Arms") category.

concentrated on two-arm spirals; we are uncertain as to whether humans are more accurately detecting spiral arms, or if they have one or more biases toward perceiving two-arm spirals. It is also interesting that the concentration in human maximum-votes for two armed spirals becomes extreme (about 97%) for the highest human agreement quartile. However, since we have a fairly large comparison sample, there are still enough galaxies to obtain meaningful distributions for the rare categories. The number of galaxies with a maximum vote for "Can't tell" also decreases dramatically when only including the highest human-agreement quartile, suggesting that arm count is not only ambiguous, but also elicits different responses in different human classifiers (instead of agreeing that the arm count is ambiguous, they distribute votes across different arm count categories).

It is likely that a better correspondence with Galaxy Zoo 2 classifications could be obtained by training an automated classifier to reproduce these classifications. The simple addition of a bias toward 2-arm classifications would likely be very helpful for classification accuracy, and automated classifiers could include more information from our output (such as arm brightness) that humans may use when counting arms. The classifier could also consist of several one-versus all sub-classifiers (one for each arm count category) that detect patterns characteristic of each arm count, or the classifier could learn a ranking function in order to determine, using our arm-level information, which arms are likely to be counted by a human. However, even with a simple arc-length threshold, we find reasonable correspondence with Galaxy Zoo classifications, suggesting that information from our output promises to be very useful for reproducing human arm count classifications, even though a simple measure leaves room for improvement as is. Since producing an unbiased description of spiral arm segment structure (from which arm counts and other measures can be calculated using the definition needed at the time) is different from finding a better mapping from our structure descriptions to Galaxy Zoo 2 arm count classifications (and thus encoding a definition of what the human classifiers consider to be an arm), we leave the latter for future work.

Chapter 6

Sensitivity Analysis

We want the output of our method to vary with the structure of the input galaxy, and to be as invariant as possible to non-meaningful differences in the galaxy images. In addition, although we expect our output to change when modifying algorithm parameters that manage structure-extraction tradeoffs (the parameters would not be useful otherwise), it is important to understand which aspects of our output are and are not affected, and to understand the degree and manner in which the affected output characteristics change. In this chapter, we explore how our output is affected (and unaffected) by image mirroring and by changes to algorithm parameters.

6.1 Image Mirroring

In human classifications of winding direction, it has been important to check for biases by mirroring some of the images shown to the humans [34, 42]. When an image is mirrored, the apparent winding direction should be reversed. In cases where the winding direction is not reversed, if one winding direction is disproportionately retained to a significant extent, then

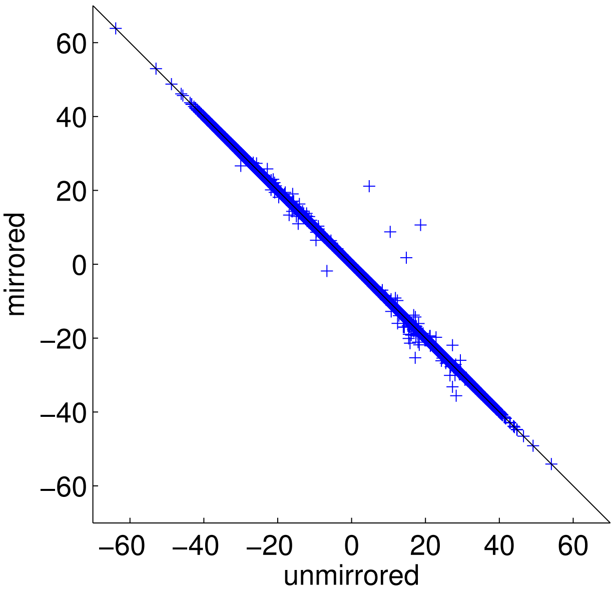


Figure 6.1: Galaxy-level pitch angles reported by our method using unmirrored and left-to-right mirrored input images across the Galaxy Zoo comparison set used in Chapter 5. These galaxy-level pitch angles are calculated as the arc-length-weighted average of all arcs agreeing with the dominant winding direction (as determined by an arc-length-weighted vote). The diagonal line gives $y = -x$. Cases on this line are visually underrepresented due to overlap.

humans likely have a bias toward that winding direction (due to human perception, the layout of the classification interface, or some other factor). As a fully computerized process, our method is not susceptible to the many potential biases in *human* perception. Nevertheless, it is useful to perform an image-mirroring test on our method. A computerized method could still have a bias through some other means (even if this is considered less likely than with humans), and an image-mirroring test provides a rare case where the exact expected output of our method is known. Furthermore, this test has allowed us to detect small imprecisions in our implementation that occasionally could affect our output (see below for further discussion), and has provided a means to verify that these imprecisions have been removed.

Figure 6.1 displays pitch angles determined by our implementation of the methods described in Chapter 3, for unmirrored images and for the same images mirrored left to right. The image set consists of the 29,250 images used for Galaxy Zoo comparisons throughout Chapter

5. The pitch angle measure is the arc-length-weighted average of all arcs agreeing with the arc-length-weighted dominant winding direction. This measure does not always consider all arcs in the image, but it does (by definition) consider most of the detected arcs by length, and we perform the flip test across a fairly large image sample (that results in a large number of detected arcs). Additionally, this measure simultaneously looks for changes to our measured winding direction and pitch angle,[1] and was used in Sections 5.1 and 5.2 when comparing with other measures of winding direction and arm tightness.

As seen from Figure 6.1, when images are mirrored, the vast majority of galaxies are given the same pitch angle magnitude (within 10^{-4} degrees) as the unmirrored case, with the sign properly flipped. Specifically, this occurred for 29,094 out of the 29,231 testable galaxies (our method did not find any sufficiently-large arm segments (pixel clusters) in the other 19 galaxies, and this outcome was consistent with respect to whether these galaxies were mirrored or not). Although reasonable to expect from an automated procedure, such a high rate of exact correspondence is more demanding than it may originally appear because any inconsistency or imprecision in any calculation for any pixel or cluster can affect the result, even if only slightly. Of the 137 galaxies where the pitch angle differed by more than 10^{-4} degrees after an image flip, all but 5 reversed winding direction, so winding directions almost always correctly flip when the image is flipped (exceptions occur for less than 0.02% of the 29,231 galaxies). There were 4 cases where the winding direction remained S-wise despite the image flip, and 1 case where the winding direction remained Z-wise despite the image flip. These results are consistent with each winding direction having equal probability of remaining through an image flip. Furthermore, in earlier versions of the code (where there were more cases of the measured winding direction "sticking," and thus a larger sample from which to test a balance between winding directions), the number of retained S-wise and Z-wise winding directions was also about equal (sometimes exactly equal). Considering

---

[1] If the winding direction changes, the magnitude of the pitch angle may change dramatically because the pitch angle measure would use the subset of arcs with the other winding direction.

the extremely low prevalence of winding direction reversal failures in the current version of our code, and the approximately balanced distribution of these failures, we find little or no evidence of a preferred winding direction in our code.

We are uncertain as to the cause of the 137 cases (about 0.47% of our sample) where there was at least a 10^{-4} degree difference in measured pitch angle. Based on previous experience, it could be caused by an imprecision or inconsistency in our code, but it may be due to equality between two or more pixel similarities (Equation 3.18). Compared to earlier versions of our code, the largest reduction in galaxy-mirroring discrepancies came from tracking the fitted galaxy center exactly during image standardization (Section 3.3); we had previously tracked it only to the nearest pixel. Since logarithmic 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. This affects the measured pitch angle. In the case of low-pitch-angle arcs, the discrepancy in the image center can spuriously flip the winding direction of the arc. If the total length of such arcs in an image is large enough, such arcs flip the measured winding direction of the image.

Other implementation changes producing improved consistency across image flips affected handling of clusters with an angular span of 360 degrees or more; if polar-coordinate θ-value disambiguation fails, the result depends on the pixel(s) that the disambiguation began with, and the selection of these starting pixels can depend on pixel ordering (as reversed by an image flip). One implementation change fixed a rare bug that arose due to an interaction of pixel quantization and certain micro-structures in the cluster boundaries (if these micro-structures were in exactly the wrong place). The other case was the annular-cluster scenario discussed in Section 3.7.2; the flip test aided the realization that such clusters do not have well-defined endpoints. The flip test also led to a fix to a rare bug in the clustering procedure where pixel similarities invalidated by a previous merge were occasionally not marked as invalid.

The remaining discrepancies may be due to a similar implementation issue,[2] but these discrepancies may be an inherent artifact of the clustering procedure itself. If two pixel similarities (and later, two cluster similarities) are the same, as calculated from the orientation field via Equation 3.18, then pixel ordering implicitly breaks the tie. If one similarity is between clusters A and B, and the other similarity is between clusters A and C (with none of the clusters equal to one of the others) then an image flip can affect whether A first merges with B or C. If the fit-based merge checking allows the first merge but not the second, then the cluster pattern can diverge from this point onward. Since pixel similarities are floating point numbers, we expect ties to be rare, but they are still possible, especially due to the sheer number of pixels (and pixel similarities). Additionally, pixel similarities only occupy a fraction of the possible floating-point range (especially due to the brightness transformation step from Section 3.2, which reduces variation in brightness and thus orientation vector magnitude), and if the unsharp mask causes image clipping, variance is reduced in the affected orientation vectors (and thus pixel similarities). Regardless of the mechanism for the flip test discrepancies, these cases are very rare, and do not show a sign of a winding direction bias.

6.2 Algorithm Parameter Changes

The algorithmic parameters discussed in Chapter 3 manage tradeoffs encountered during spiral galaxy structure extraction. Changing these parameters will, by their nature, at least slightly alter the behavior of our method. Here, we characterize the effects on our output when varying six algorithmic parameters at the core of our procedure. For each such parameter, we run our method on our Galaxy Zoo comparison set (used in Chapter 5 and Section 6.1) with five alternate parameter values.[3] If output could not be produced

---

[2] Due to time constraints, this issue could not be completely eliminated.

[3] Due to CPU time constraints, we present results from a slightly old version of our code. The main change relevant here is that the default unsharp mask was 10 instead of 6, but other than the difference in

for at least one of these parameter values, the galaxy was dropped from the analysis for that parameter.[4] We then compare, on a per-galaxy basis, changes to several aspects of our output. Specifically, we examine effects on the fragmentation and volume of arcs detected by our method, as measured by the average arc length, total arc length, and total number of arcs. We also measure effects on our detected winding direction and pitch angle. For the pitch angle we again use the arc-length-weighted average of all arcs agreeing with the arc-length-weighted winding direction, referred to here as the "signed pitch angle" (the sign determines winding direction). To isolate the effect on the measured arm tightness, we also measure changes in the absolute value of the pitch angle (higher absolute values indicate looser arms). Lastly, we also consider changes to winding direction. Winding direction is expressed as 1 for S-wise, -1 for Z-wise, and 0 for no winding direction (a galaxy-level pitch angle of exactly zero). Consequently, a S-to-Z winding direction change has a winding direction difference of -2 and a Z-to-S winding direction change has a winding direction difference of 2. Similarly, changes to or from a non-winding-direction value of 0 are expressed as differences of -1 or 1, but this type of change is rare because almost all galaxies have a measured winding direction.

Figure 6.2 illustrates the effects of varying the minimum cluster size (in pixels) needed for a cluster (and its associated arc) to be included in our output. This parameter was discussed in Section 3.7.2. The red lines give the median changes (in the measured aspects of our output) resulting from each setting of this cluster size threshold (relative to the baseline value of 150 used elsewhere in this work). The red error bars give the upper and lower quartiles of this difference. Similarly, the green error bars give the 10th and 90th percentiles of this difference, and the blue error bars give the 5th and 95th percentiles. Unsurprisingly, the average arc length increases to the extent that we require larger cluster sizes. The typical

---

this default, we do not expect the minor code changes to substantially affect our results, especially since we saw no substantial difference when performing these tests on a version of the code earlier than the one used here.

[4] The number of affected galaxies was too small to substantially affect our plots, except in the case of the unsharp mask amount. We investigate this potential change when examining the effect of this parameter.

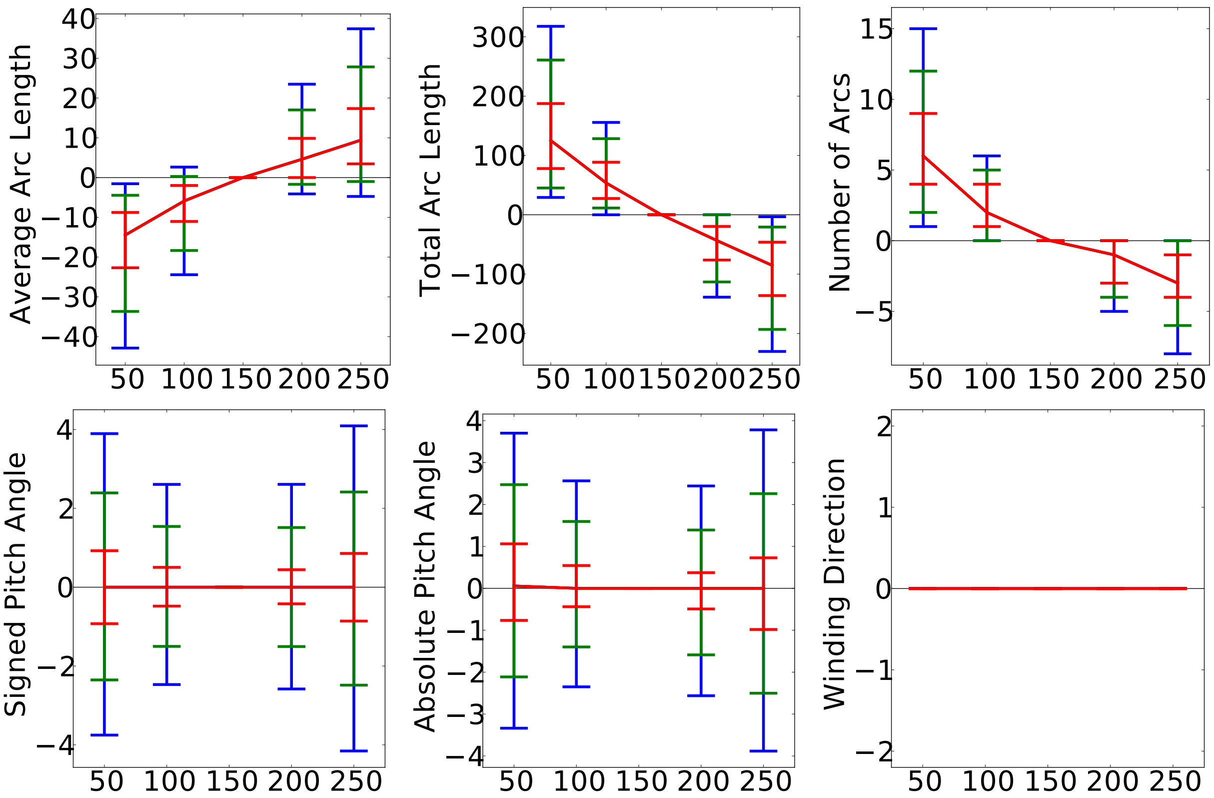


Figure 6.2: Changes in several measures computed from our output when varying the minimum size needed for a cluster to be included in our output (x axis). The median change is given as the red line, with the red error bars giving the upper and lower quartiles. The green error bars give the 10^{th} and 90^{th} percentiles of the changes to our output, and the blue error bars give the 5^{th} and 95^{th} percentiles.

change in average arc length is about 10, but this difference can increase to about 40 in some cases. Allowing fewer small clusters into the image also decreases the total arc length, as expected. It is also obvious that a less permissive size threshold will reduce the number of clusters; the degrees of change and variation indicate how many clusters tend to be discarded in this manner. Finally, we note that changes to the measured pitch angle are minor, with no (or negligible) change to the typical pitch angle measured, and fairly low scatter. The use of arc-length weighting likely increases the stability of the measured pitch angle, but all clusters (that agree with the dominant winding direction) nevertheless contribute to the pitch angle measurements. Winding direction also tends to be unaffected; from the figure (in particular, the absence of visible blue bars) we note that at least 90% of the images (and

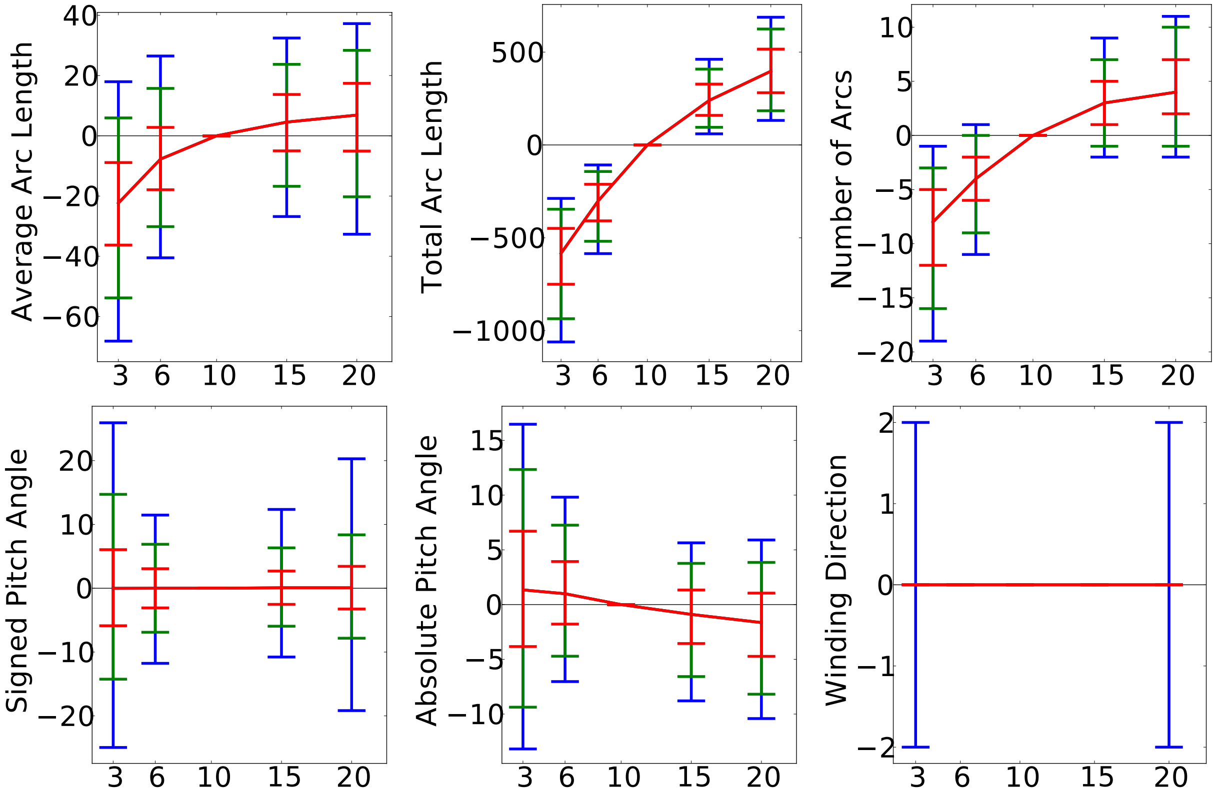


Figure 6.3: Changes in several measures computed from our output when varying the unsharp mask amount (x axis). The median change is given as the red line, with the red error bars giving the upper and lower quartiles. The green error bars give the 10th and 90th percentiles of the changes to our output, and the blue error bars give the 5th and 95th percentiles.

probably more) are unaffected in winding direction.

Figure 6.3 is set up in the same way as Figure 6.2, but instead examines the effects of varying the unsharp mask amount (Section 3.5.2).[5] Increasing the unsharp mask amount increases the contrast between bright and dark regions, increasing the magnitudes of arm-aligned orientation vectors and thus increasing the pixel similarity scores within spiral arms (since the similarity scores are computed by a dot product; see Equation 3.18). This, in turn, increases the number of pixel and cluster merges, which means that clusters tend to be larger (more pixels merge into the clusters) and have a higher tendency to merge (since clusters are more likely to be adjacent to each other when they grow). Since larger

---

[5] As discussed earlier, the baseline unsharp mask amount is 10 instead of 6 because we use a slightly old version of our code.

clusters are more likely to reach the minimum size threshold, and since clusters are more likely to be picked up by the orientation field when the unsharp mask is stronger (up to a point), we see an increase in the total number of arcs. This trend overpowers the reduction in cluster count caused by an increased number of merges, but the cluster-count increase slows for larger unsharp mask amounts, suggesting that large-cluster merges accelerate for larger unsharp mask amounts, or that incremental increases in the unsharp mask amount have a stronger impact on orientation field sensitivity when increasing from a lower value of the unsharp mask (both of these potential effects may play a role). The higher propensity for cluster merges increases the average arc length, in spite of the decrease in this length induced by the introduction of smaller clusters that meet the size threshold under a stronger unsharp mask. The slowdown in the increase in average arc length could indicate that larger numbers of smaller clusters are being introduced for higher unsharp mask amounts, or that cluster growth becomes more width-wise for larger unsharp mask amounts. The total arc length increases strongly and steadily for all increases in the unsharp mask, due to both the introduction of new clusters and the growth of existing clusters.

Changes in the unsharp mask amount do not change the typical signed pitch angle measured by our method, although the variance in pitch angle changes is relatively high. The measured arm tightness does increase slightly (the absolute pitch angle decreases slightly) with a stronger unsharp mask. The median change in typical arm tightness is relatively small (about 0.99 and 1.34 degrees with unsharp mask amount decreases, and about 0.90 and 1.64 degrees with unsharp amount increases), but since it varies smoothly with unsharp mask amount, the change in tightness might be a real effect. We are uncertain as to the cause, but possibilities include: slightly looser arms (higher absolute pitch angles) at the brightest parts of the spiral arms (which do not need a stronger unsharp mask to be picked up by the orientation field); the possibility that “real” spiral arms are, on average, slightly looser than noisy arcs (that are more prevalent with a stronger unsharp mask); a (slightly) elevated risk of detecting the edge of the galaxy disk as a low-pitch-angle spiral arm (or increased sensitivity to legitimate ring patterns at the edge of the galaxy disk) when the unsharp mask amount is increased; a slightly increased tendency to merge very-low-pitch-angle arcs into a zero-pitch-angle ring when the unsharp mask amount is increased; a slight tendency for pitch angle to be reduced when clusters are merged (but our later analysis of cluster-merge leniency does not find this to be the case); and a possible tendency for brightness gaps to have a sharper brightness decrease (to the point where the gap is exaggerated instead of reduced by the unsharp mask) for arms with a higher pitch angle (e.g., arms with a high (or at least not-extremely-low) pitch angle may have a higher chance of containing a severely-interfering dust lane, in which case a stronger unsharp mask is more likely to fill gaps in low-pitch-angle arcs). Winding direction is mostly unaffected, although there are some flips in both directions for the severe unsharp mask values, since the unsharp mask can strongly affect the prominence and inclusion of clusters (and thus the presence and strength of winding direction votes).

At the lowest tested unsharp mask amount, the number of images where output could not be produced (due to a lack of sufficiently-sized clusters) was much higher than normal, requiring us to discard many more galaxies than normal (7094 out of 29250). If we remove the lowest unsharp mask amount value from the analysis (so that the number of unavailable galaxies is reduced to 755, which is still relatively high, but now only about 2.6% of the galaxies), the results are almost the same, except that winding direction flips are visible in the “blue” error bars (5% and 95% difference percentiles) for an unsharp mask amount of 15.

Next, we assess the impact of the scale of the unsharp mask (Section 3.5.2); the effects of changes to this parameter are shown in Figure 6.4. The scale of the unsharp mask controls the width of the Gaussian blur subtracted from the image. Consequently, smaller scales emphasize contrast in smaller image features and larger scales emphasize contrast in larger image features. The effect of the unsharp mask scale on the total number of arcs is unique in that it is not monotonic; typical arc counts decrease with either an increase or decrease

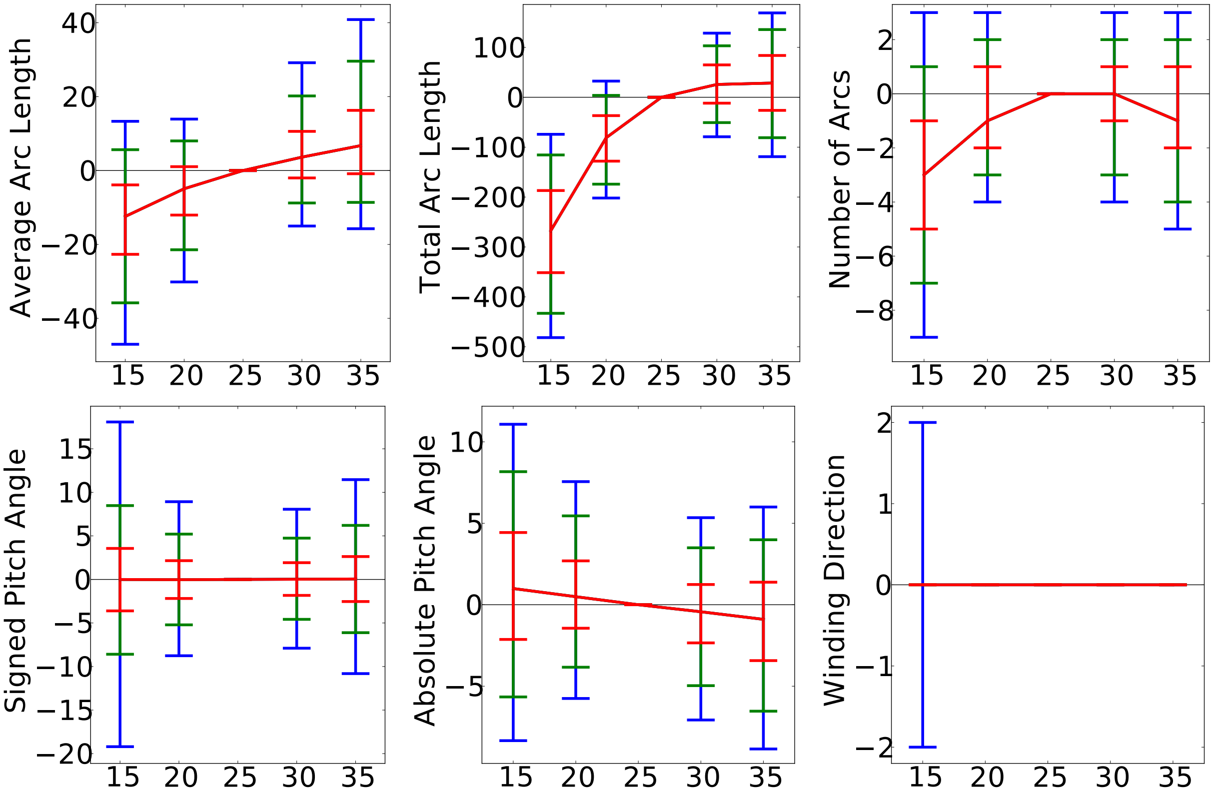


Figure 6.4: Changes in several measures computed from our output when varying the size scale of the unsharp mask (i.e., the scale of the Gaussian blur subtracted from the image) (x axis). The median change is given as the red line, with the red error bars giving the upper and lower quartiles. The green error bars give the 10^{th} and 90^{th} percentiles of the changes to our output, and the blue error bars give the 5^{th} and 95^{th} percentiles.

to the unsharp mask scale. This confirms that the unsharp mask scale is within a range that emphasizes the typical size scales (widths) of spiral arms found in the standardized images. Increasing the size scale of the unsharp mask increases the average arc length, perhaps because larger-scale features are emphasized. The total arc length increases as a function of the unsharp mask scale, but less so at the largest unsharp mask scales, perhaps because the arc count and average arc length both decrease at unsharp mask scales smaller than the baseline value of 25, but move in opposite directions at unsharp mask scales larger than the baseline.

There is no substantial difference in the typical signed pitch angle, but the arm tightness increases slightly (the absolute pitch angle decreases slightly) with larger values for the unsharp mask scale. The effect seen is very similar to the effect of increasing the unsharp mask amount, so the underlying mechanism (or mechanisms) may be similar. We also note that decreasing the unsharp mask amount and decreasing the unsharp mask scale may both favor smaller portions of the spiral arms (due to the smaller portions' higher contrast for the unsharp mask amount and their smaller size for the unsharp mask scale). Additionally, both algorithm parameters would be affected by the severity of brightness decreases between arm segments that could plausibly be merged together; as discussed previously, this severity could vary slightly by pitch angle, on average. These observations further support the hypothesis that a common mechanism is present for the (slight) decrease in absolute pitch angle found when increasing the unsharp mask amount or unsharp mask scale. The effect of the unsharp mask scale on typical winding direction is simpler: typical winding directions are unaffected, but some flips are noticeable at the smallest tested unsharp mask scale, likely because noise is more prevalent at smaller scales.

Figure 6.5 displays the effect of varying the maximum allowed combined-cluster to separate-cluster error ratio (i.e., Equation 3.19 in Section 3.7.2), which controls the stringency applied when checking cluster merges (higher values are less stringent). As the allowed ratio increases, the average arc length increases; this is expected because clusters get larger when more merges are allowed. Reduced merge stringency also decreases the total number of arcs. This parameter change could also conceivably increase the number of arcs because increased merge-check leniency could allow clusters to grow beyond the size-based output-inclusion threshold, but since fit-based merge checks probably have a stronger tendency to block merges of large clusters (there are more ways in which large clusters can deviate from the logarithmic spiral model), it makes sense that the net effect of less stringent checks is to reduce the number of arcs (since an arc merge decreases the count by one). Less stringent checks also reduce the total arc length, although the impact of this parameter change on total arc length is less than the effect of most other algorithm parameters. The reduction in total length is sensible because the merged arcs can sometimes overlap slightly in their

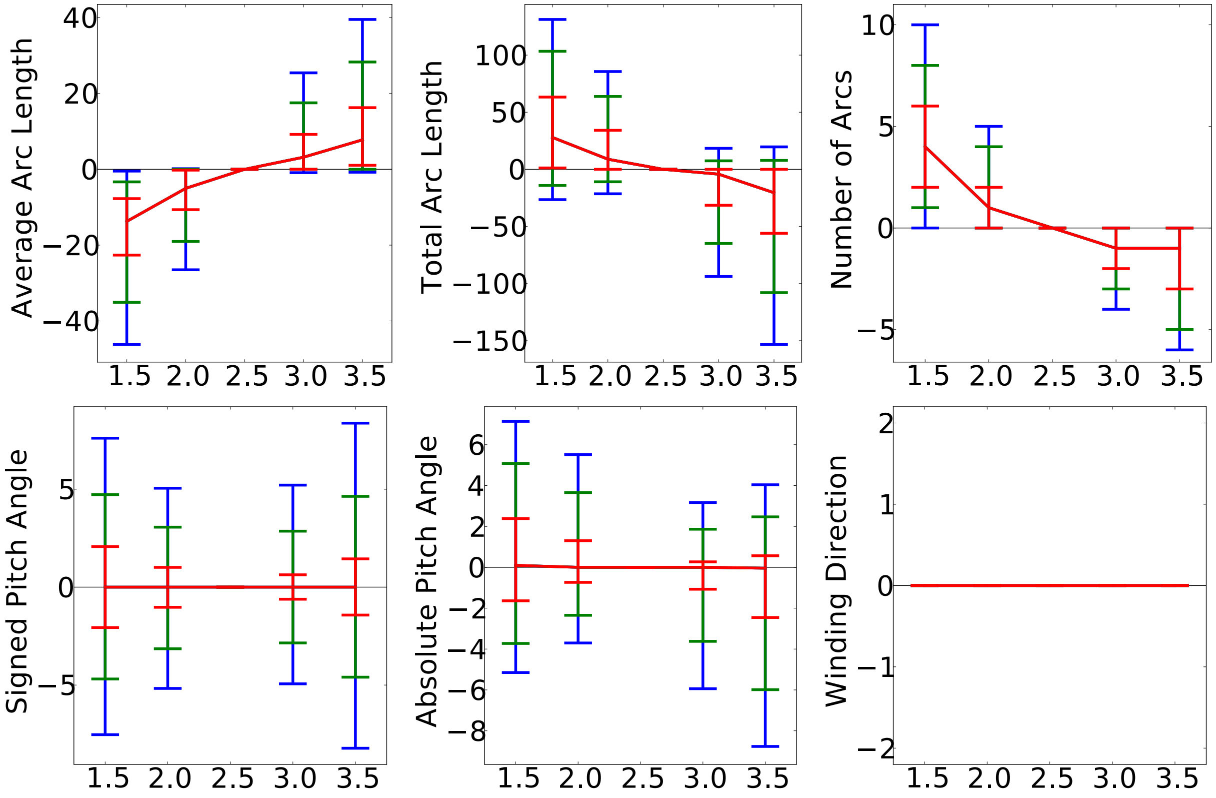


Figure 6.5: Changes in several measures computed from our output when varying the maximum allowed combined-cluster to separate-cluster error ratio (x axis). The median change is given as the red line, with the red error bars giving the upper and lower quartiles. The green error bars give the 10th and 90th percentiles of the changes to our output, and the blue error bars give the 5th and 95th percentiles.

angular range. However, it is interesting that (for most galaxies) this effect is stronger than the extra arc length gained when allowing more merges of non-contiguous clusters during the secondary merging step (Section 3.7.2). The effects on the typical signed pitch angle and arm tightness (absolute pitch angle) are negligible to nonexistent, suggesting that difficult merge decisions do not, on average, change the typical pitch angle. Similarly, effects on winding direction are rare or nonexistent.

As discussed in Section 3.7.2, this fit-based merge check is only used when both clusters reach a minimum size. Figure 6.6 displays the effects of varying this minimum. For all of the output measures considered, typical values do not change substantially; to the minimial extent that values do change (examining the heights and asymmetries of the bars), the

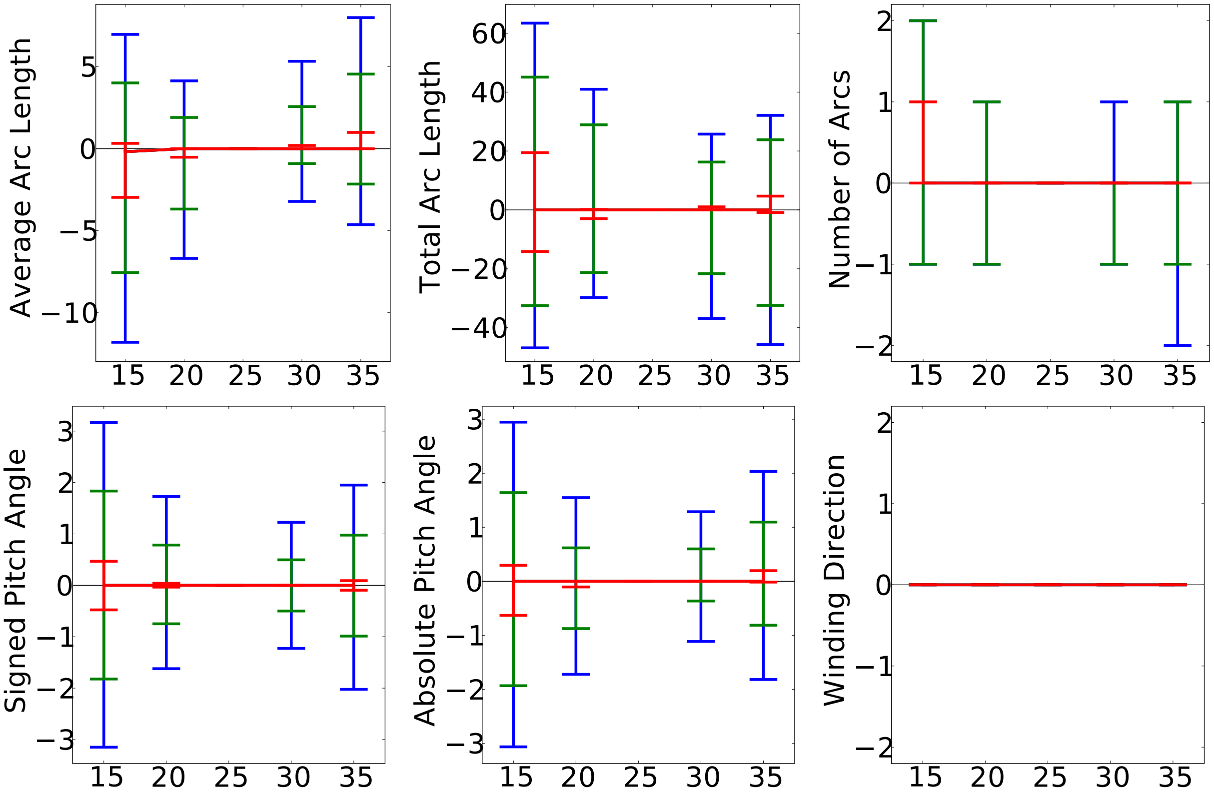


Figure 6.6: Changes in several measures computed from our output when varying the minimum size each cluster must reach in order to trigger a fit-based merge check (x axis). The median change is given as the red line, with the red error bars giving the upper and lower quartiles. The green error bars give the 10th and 90th percentiles of the changes to our output, and the blue error bars give the 5th and 95th percentiles.

effects of increasing this merge-check size threshold are similar to the effects of making the fit-based merge checks more lenient. This is sensible because increasing the merge-check size threshold also makes the merge checks more lenient, but in a more limited way: checks are more important for large clusters, but this minimum is only relevant for very small clusters. Since this minimum size was used in part to save computation time, it is good that changes to its value do not result in any substantial changes to typical values of any of the measurements. The other aim of this minimum was to avoid performing a merge check before the cluster shape was known. The smallest tested minimum value (which starts merge checks earlier) produces the most scatter, suggesting that this objective is being met. We also note that scatter begins to increase again at the largest tested minimum value, which further supports

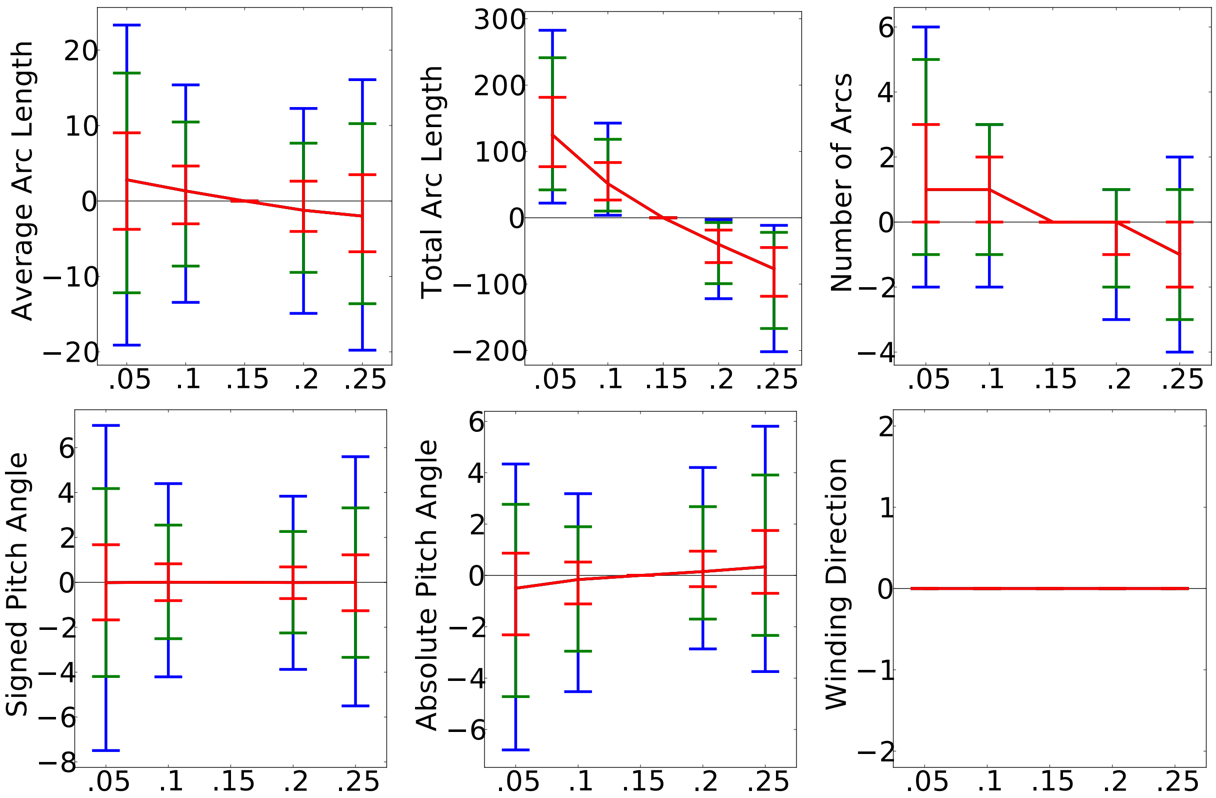


Figure 6.7: Changes in several measures computed from our output when varying the minimum orientation similarity needed to continue clustering (x axis). The median change is given as the red line, with the red error bars giving the upper and lower quartiles. The green error bars give the 10th and 90th percentiles of the changes to our output, and the blue error bars give the 5th and 95th percentiles.

the conclusion that the current parameter setting is a good one; the increased scatter may be due to allowing clusters to grow too far before starting merge checks. In any case, however, there appears to be little or no effect on pitch angle, tightness (absolute pitch angle), or winding direction.

Finally, we investigate the effect of the stopping threshold applied to the pixel clustering (recall from Section 3.7.2 that cluster similarities are determined by the maximum similarity of inter-cluster pixel pairs, that pixel similarities are determined by the dot product of the associated orientation vectors, and clustering continues in similarity order until the next cluster similarity falls below the threshold value considered here). Comparing to Figure 6.3, which characterized the effect of the unsharp mask amount, we see that all trends are reversed, but less strong in terms of both the amount of scatter and the change in median values. This correspondence in effect is sensible because both the clustering stopping threshold and the unsharp mask amount control the extent to which clusters can expand (with more expansion for a higher unsharp mask amount or a lower stopping threshold); higher unsharp mask intensities increase many “good” cluster-to-cluster similarities, while reducing the stopping threshold makes the pixel similarities larger relative to the stopping threshold. Compared to changes in the unsharp mask amount, changes in the stopping threshold may have a weaker effect due to the range of stopping threshold values tested here, but it may also be because the stopping threshold uses more nuanced information (the next merge depends on cluster merge history, where the best merges were tried first) and is moderated by the fit-based merge checking. Examining the effects of the stopping threshold on the individual measurements and keeping in mind that lower thresholds mean more cluster growth, it is sensible that a lower stopping threshold slightly increases average arc length (since existing clusters can grow by incorporating boundary pixels, and since merges reduce the cluster count, with these two factors overpowering the introduction of small clusters that newly exceed the size threshold), total arc length (as a straightforward consequence of cluster growth, along with a greater chance for smaller clusters to reach the minimum size threshold), and (to a slight extent) the total number of arcs (suggesting that, on average, slightly more clusters are introduced by exceeding a size threshold than are removed due to merging). As was seen with the unsharp mask intensity, there is no noticeable effect on the typical signed pitch angle, but the arm tightness is increased somewhat (the absolute pitch angle is decreased somewhat) with more cluster growth. Since this (weak) trend is similar to the observed effect of increasing the unsharp mask amount, and due to the previously-mentioned similarities of the stopping threshold and the unsharp mask amount, the underlying mechanism behind the arm-tightness change is likely similar. Lastly, we find that typical winding direction is unaffected, as usual.

Chapter 7

Future Directions

7.1 Modifications to the Current Method

The spiral galaxy structure-extraction method described in Chapter 3 is capable of producing useful information about any kind of spiral arm-segment structure, and has been shown to agree well with previous measurements where such measurements are available (see Chapter 5). Due to the wide variation in spiral galaxy structure, caution is advised when considering ideas for improving this method: even if a change is beneficial (or imagined to be beneficial) to some galaxies, it is common for the same change to be detrimental in other cases. Each change should thus be tested against an image set with “known” measures or classifications of spiral galaxy structure; this will help identify situations that should be examined for potential loss of accuracy, even if it is not desirable to fully tune results to match existing structure measurements and classifications. We did so in the later developmental stages of this work (where baseline accuracy was high enough to necessitate such comparisons), primarily using a set of 29,250 images with corresponding Galaxy Zoo classifications; Chapter 5 describes this set and the comparisons made with it. Nevertheless, despite the possibility that a modification will be unexpectedly counterproductive, some changes to this method could yield incremental improvements. Here, we list several possible changes that could be explored.

Star masking (Section 3.4) has been helpful in avoiding disruptions to the ellipse fit used in the image standardization procedure described in Section 3.3. Since SExtractor (the program used to group pixels corresponding to the same light source [7]) does not always assign the galaxy to exactly one pixel region, the star mask is removed after the image standardization step in order to avoid the risk of removing part of the galaxy before spiral arm-segment detection. However, star masking could still be useful after image standardization; when foreground stars (or other objects) are close to the galaxy (close enough to be included in the standardized image), they can produce spurious clusters if their brightness patterns are picked up by the orientation field. It might be possible to use SExtractor in a more sophisticated manner in order to more reliably determine which SExtractor objects are part of the target galaxy and which are not. Potential ways to do so include training the SExtractor star/galaxy classifier for the image set used (or building one's own classifier), or tuning the SExtractor parameters for the dataset of interest. Alternately, one could address some problematic cases by defining heuristics that identify SExtractor regions that are almost certainly not part of the galaxy (possible criteria include not being part of the primary or extended regions discussed in Section 3.4, and/or having a large distance from the image center). Once non-galaxy regions are known and a corresponding mask is produced, the masked regions could be mapped to their locations in the standardized image (with nearest-neighbor interpolation, rounding, weighting, or some other way of handling the lack of one-to-one correspondence between original and standardized-image pixels), and then the mask could be used to reduce or remove orientation vectors in the masked regions (it is probably best to avoid applying the non-galaxy-region mask before the unsharp mask (Section 3.5.2), in order to prevent unsharp mask artifacts). Removing these orientation vectors would prevent creation of spurious clusters that may otherwise arise from objects other than the target galaxy.

SExtractor output could also be used to identify cases where the ellipse fit (used in image standardization) may have “zoomed in” too far and cut off dim arms. This is rare, but as mentioned in Section 5.1, it does happen occasionally. Since at least one of the SExtractor regions (usually) corresponds to the galaxy image, the ellipse fit is suspicious if its contour is too far inside the SExtractor region (or regions) corresponding to the galaxy. If such a condition arises, a warning could be given, or (if confidence in the SExtractor region is high enough) the ellipse fit could be stopped at the iteration before it shrank too far within the SExtractor galaxy region (or regions). However, the success of such changes would depend on the reliability of SExtractor output (or information computed from it).

If one assumes that input will always be available as FITS images, one could apply separate brightness transformations (Section 3.2) for the image standardization step and the unsharp mask step, since the ideal brightness-transformation parameters may not be the same for both purposes. Specifically, the brightness transformation used for the image standardization step would not need to preserve arm-disk contrast (it would only need to preserve the contrast between the disk and the sky). By more aggressively increasing the brightness of inter-arm regions, unusually dim arms would also be brightened, reducing the risk of cutting off low-brightness arms. After using one brightness transformation to compute the ellipse parameters for image standardization, the brightness transformation could be changed before the actual image standardization (de-projection) operations.

In Section 3.6 we described a method for reducing interference from bars. Galaxy bulges can also produce unwanted orientation vectors, and can thus interfere with pixel clustering. Many galaxies do not have a bulge prominent enough for this to occur, and when there is a prominent bulge, its cluster is often removed by the center-cluster deletion. However, in the remaining cases, small low-pitch-angle arcs can be present near the galaxy bulge, or in especially severe cases, an arm cluster can be merged with the bulge cluster, causing the arm cluster to be removed along with the (center-containing) bulge cluster it merged with. In response, a bulge model (e.g., a centered ellipse that accounts for the fact that the spherical bulge can appear ellipsoidal after the image standardization step) could be developed and used in the same way as the bar model (i.e., the bulge model would replace a logarithmic spiral arc if the former is a better fit to the cluster). Alternately, the bulge region could be removed from the image or the orientation field, if the bulge region is known. As a crude way to find the bulge region, the ellipse-fitting procedure (and its stopping criterion) from the image standardization step (Section 3.3) could be continued past the point where the galaxy disk outline is detected; the disk-fit stopping condition usually applies again when the ellipse fits the bulge (when the condition is never triggered, the bulge is usually far too small to be a concern for our clustering). The bulge region could then be traced through the image standardization (de-projection) operations, and orientation vectors corresponding to this region could be reduced or removed. An option to approximately fit the bulge already exists in our code,[1] and could be applied for this purpose (at the expense of increased computation time, especially for high-resolution input images). Alternately, the bulge (and possibly also the disk and bar) could be fitted using bulge-disk decomposition software (see Section 2.2 for some possibilities), and then subtracted from the image.

Fit-based merge checking (Section 3.7.2) usually does well in ensuring that clusters do not encompass more than one spiral arm segment. However, since the merge criterion (Equation 3.19) is relative to the individual fit errors of the clusters involved in the proposed merge, it is (at least in theory) possible for "runaway" merges to occur, where cluster fit error becomes bad enough for almost any merge to be acceptable, and where such merges make even worse merges acceptable, so that most or all contiguous arm segments are contained in the same cluster, even if this cluster contains multiple arms. After reweighting the fit to the combined cluster (during the merge check only) so that both clusters in the proposed merge have equal

---

[1] This fit is not intended to be precise enough to have a physical interpretation; if such information is desired, it is still necessary to use a bulge-disk decomposition method.

weight (see Section 3.7.2), we have not encountered this scenario, in large part because fit error increases are usually severe when attempting to combine two distinct spiral arms into one cluster (especially due to the use of the squared error, as discussed in Section 3.7.1), with the weight balancing ensuring that clusters do not gradually accumulate poorly fitting regions. Additionally, since Equation 3.19 uses the larger of the two individual-cluster error increases, the runaway merge conditions (bad fit error before merging) must apply to both clusters separately. However, if runaway merges occur after future changes, or if it becomes desirable to apply additional cluster-quality conditions, one may wish to impose a maximum fit error (per unit arc length or per pixel) by blocking merges that would exceed this limit. The implementation of this condition would be simple, but one would need to be cautious of its consequences on the clustering, especially since an absolute threshold would increase reliance on the logarithmic spiral model.

If two low-pitch-angle arcs could conceivably be merged into an annular cluster (that may or may not be a true ring), either a two-arm interpretation or a ring interpretation could be valid (including in terms of the fit error). Since the two-armed interpretation occurs first (clusters develop by merging), the two arcs would be merged into an annular cluster if the resulting fit is acceptable. As described in Section 3.7.2, these clusters are given a pitch angle of zero because arc endpoints are no longer well-defined. Figure 5.3 includes an example of this scenario. Although the currently-allowed cluster merges are almost always informative (a single spiral arm segment is traceable for a longer extent, and the arm segments are more coherent), one may decide that, in the special case of merges into annular clusters, there is a substantial risk of losing information (for example, the cluster does not have a winding direction because the pitch angle is zero). In this case, it is reasonable to consider penalizing (but not completely preventing) merges into annular clusters (for example, by using a stricter merge-check threshold for this case). Such a penalty could reduce merges of low-pitch angle arcs (the fit error would increase substantially where the two arms stick out at their ends), while still allowing rings to pass the merge criterion. However, the existing merge criterion may already be sufficient in almost all cases, so one would need to determine if there are enough problematic cases to warrant an attempted change, and make sure that such a change does not inadvertently interfere with cluster quality in other ways.

In some galaxies, there may be more than one way to merge clusters into a plausible structure description (for example, when arms have a fork, the "handle" can often be continued into either of the prongs). In such cases, it may be useful to detect ambiguous merges, notify the user that such a merge occurred, and possibly also provide other plausible merge sequences. This would be possible to some extent by maintaining a merge history for each cluster, and considering similarities and merges with these earlier clusters in addition to the current ones. Doing so would require additional computation time, which could be controlled by limiting the length of the merge history.

For many (perhaps all) galaxies, it would be useful to develop a measure of cluster quality. Such a measure could be used as a weighting scheme more sophisticated than arc length alone, and could help flag galaxies where manual review could be useful. Fit error would also be a useful signal here, but since small noise-induced clusters (or other spurious clusters) can sometimes have good fits, fit error does not always distinguish between "real" arms and spurious clusters, especially for small arcs (where this distinction may be needed the most). For longer arcs, somewhat-higher fit errors may be due to variations in arm width or due to small deviations from the logarithmic spiral model (cluster merges allow some deviation), rather than uncertainties about the arm or cluster itself.[2] In addition to fit error and arc length, other useful signals could include cluster brightness, arm-interarm contrast, and the degree of pixel-level brightness variation relative to the total brightness range (perhaps within the input image). This cluster-quality measure would need to be developed with caution because many attributes are not as discriminative as they may first appear (for example, noise-induced clusters are usually dimmer, but low-surface-brightness arms can

---

[2] Cluster merging is much more resistant to this issue, since we only need to compare relative errors (of merged and unmerged clusters), rather than producing an absolute measure (of cluster confidence).

also be dim, and arm segments containing knots of bright stars may not necessarily be more reliable than arm segments without these bright stars). It may be helpful to use supervised machine learning techniques to find such a confidence measure; for example, one could use the attributes of the arc-level CSV (Section 4.2) to try to predict whether an arc agrees with the "correct" winding direction (as determined by our arc-length weighted vote, by Galaxy Zoo, or by some other source).

Our spiral arm structure information is given in terms of a standardized image; if desired, the ellipse-fit information we provide (Section 4.3) can be used to translate this information to the view given in the original image.[3] However, in some special cases, it may be useful to perform our structure-extraction procedure directly on the given galaxy image, which could have a resolution substantially different than the 256 × 256 size used throughout this work. This will likely require adjustment of algorithm parameters that depend on the image resolution. In particular, it might be useful to increase the number of orientation field levels (Section 3.5) by one for each doubling in image resolution (so that the resolution of the coarsest orientation field level is approximately consistent), and the maximum cluster distance allowed in secondary merging (Section 3.7.2) should probably increase linearly with the (linear) image dimensions. It may also be useful to adjust the minimum cluster size in proportion to the number of pixels in the image. Additionally, since the spiral arm widths (in units of pixels) may no longer be within a constant range if the image standardization step is bypassed completely, it would be necessary to adjust the unsharp mask scale on a per-image basis. This could be done automatically, perhaps by performing the ellipse fit that would have been used for image standardization, but using this ellipse fit to adjust the unsharp mask scale rather than to de-project the image. The threshold on the fit error ratio used in fit-based merge checks (Equation 3.19) would probably also require adjustment, since squared error would increase quadratically with linear scaling in the image dimensions, making merge checks stricter (for a fixed error ratio) when the image size is increased.

---

[3] Providing this capability in our own code is also a possibility for future work.

Alternately, this arc-merge criterion (Equation 3.19) could be modified so that the square root of the error ratio is taken before applying the threshold, although some image-size-based adjustment may still be necessary. If it is worthwhile to perform spiral arm structure extraction directly on input images (even though our output information can be re-projected to the input image), the exact changes in these algorithm parameters could be determined. However, one should note that it is not necessary to do so if one only wishes to use deprojection (disk-fit) parameters from another method (e.g., a bulge-disk decomposition), as one could simply compute a bounding ellipse using this external method and then apply the rest of our image standardization procedure.

Our output assigns each pixel to exactly one arm segment (or the background). However, many pixels may, to varying extents, be associated with multiple galaxy components (and possibly also the background sky). Consequently, it may be useful to provide "soft" cluster assignments instead of associating each pixel with exactly one arm segment (or the background). The most natural way to do so would likely be with a brightness fitting method (discussed in Section 7.3). Each pixel brightness value would be modeled as the sum of brightness values from all components, and so pixel membership could be naturally expressed as the proportion of brightness contributed by each component. Alternately and more generally, any model that describes features (brightness, color, contrast, or other measures) of galaxy components could be used with an expectation maximization algorithm [17] that iteratively alternates between estimating galaxy component parameters and degree of galaxy component membership. Our output could provide a highly informative initialization for such an algorithm.

Lastly, although not an integral part of our structure-extraction procedure, somewhat application-dependent, and at least partially under the discretion of those using our code, it would be interesting to use our output to develop additional measures of galaxy structure (e.g., lopsidedness), or create more sophisticated versions of existing measures. For information and further discussion related to this goal, see Chapter 4.

7.2 Spiral Arm Segment Extraction via Scale-Space Ridge Detection

As discussed in Section 3.5, transforming pixel brightness values into orientation field vectors (based on the values of nearby pixels and as originally described by Au [2]) provides substantially more information about whether or how a pixel belongs to a spiral arm. Using this information, we were able to devise a clustering scheme that determines which pixels belong to the same spiral arm segment, and associate each cluster with a logarithmic spiral arc. The orientation field is thus quite useful, but a more general observation is that, when determining a galaxy's spiral arm "skeleton," local patterns of pixel brightness are often more informative than the pixel brightness values themselves. It is thus reasonable to consider other ways to look for these brightness patterns.

We have started building a new spiral arm structure-extraction system that also produces a list of spiral arcs and associated pixel clusters, but determines this information through different means, starting with a different way of finding brightness patterns characteristic of spiral arms. This new method handles arm-size variation more naturally, both in terms of the overall resolution of the galaxy and in terms of within-galaxy arm-width (and arm-length) variation. It also does not require pixel-level clustering; instead, the new approach to analyzing pixel brightness patterns naturally produces candidate pixel regions at different size (arm-width) scales. The candidate regions are partially hierarchical; some candidate regions contain or overlap substantially with one or more other candidate regions. Consequently, the new method still performs cluster growth and merging, but these operations are only needed when making decisions about cluster-to-arm-segment correspondence, not out of necessity in growing the clusters themselves. In addition, unused candidate regions can provide information about ambiguous region layouts (arm-segment assignments), such as decisions about which arm-fork prong to join with the "handle." Furthermore, since the set of possible regions is known at the start, our decisions can be more forward-looking while still retaining the tractability of a greedy algorithm.

In this method we begin by noting that spiral arms have a ridge pattern in their brightness: local maxima occur across the arms, and the brightness is somewhat consistent (and higher than surrounding inter-arm regions) along the arm. Lindeberg [37] provides a means to calculate pixel-level ridge strength as a function of the ridge scale. In our case, the ridge scale corresponds to the approximate width of a spiral arm, so by using multiple scales, we can detect arms of different width within each image. At the finest scale feasible under the resolution of the given image I (after a reasonable amount of up-scaling, if desired), the ridge strength of each pixel is calculated as

$$R = (\partial_{xx} I - \partial_{yy} I)^2 + 4(\partial_{xy} I)^2 \ , \tag{7.1}$$

where the partial derivatives are computed via finite differences of adjacent pixels. This measure is one of multiple ridge-strength options discussed in Lindeberg [37]; we use this one because (as discussed in Lindeberg [37]) it has the lowest response to Gaussian "blobs" (i.e., points that are local maxima in both directions), which correspond to unwanted non-arm features such as stars and the galaxy core. This ridge-strength measure responds to both positive and negative ridges (dips), but the sign can be calculated as well; see Lindeberg [37] for details. To detect ridges at scales (widths) larger than about one pixel, a Gaussian convolution (blur) can be applied to the image before computing the ridge strength; by doing this for different Gaussian scales (widths), ridge strength can be obtained as a function of both image position and ridge scale [37]. We can then perform this ridge detection on an image where the brightness has been transformed using the same methods as Section 3.2, and where the image has been de-projected (but not necessarily resized to a standard size) as described in Section 3.3.

Lindeberg [37] then finds ridge points by incorporating a normalizing factor that varies as a function of scale, and then finding local maxima in the space of image positions and scales. It is here that our method would likely diverge. We attempted to use this scale-normalized ridge strength measure but found that it was not sufficiently reliable for our purposes, perhaps due to correlations between scale and brightness found in galaxy images, interference due to ridges occurring within ridges (e.g., an arm may have a narrow band of bright stars), and interference from the galaxy core and from stars (within the target galaxy or within our own).[4] It may be possible to tune the scale normalization for our purposes, but we also want to determine the entire region corresponding to each spiral arm segment, so we envision a different route — especially since, in our case, we have another option for choosing (and combining) scales, using the fact that image regions should approximately correspond to logarithmic spiral arcs.

Our procedure uses the ridge-strength measure (Equation 7.1) to extract candidate arm regions. To do so, we first choose a set of scale values, and then compute ridge strength (as a function of image position) for each of these scale values. The range of these scale values can vary according to the image size, so that all possible arm widths can be considered. The spacing between scale values determines how densely one looks through the range of possible arm widths. In situations where a scale value is close to the widths of one or more arms in the image, we have found that ridge-strength values will be coherently high within these spiral arms. In particular and most importantly, in our experiments so far we have found that for each spiral arm with an approximately-matching scale there will be a range of ridge-strength contour values that delineate the approximate boundary of the arm. Unfortunately, like arm

---

[4] Even though these blob-like objects are supposed to be ignored by the ridge strength measure we chose, they are often not perfectly isotropic, and are often very prominent, even after brightness transformation. Consequently, they can still overpower spiral arms, especially when combining pixel-level information from different scales.

width, these contour values are not known in advance, so (also like arm width) we choose a set of contour values. To robustly choose these contour values despite potentially nonlinear, galaxy-dependent, and scale-dependent fluctuations in the ridge strengths, we choose the contours via ridge-strength percentiles. For each scale, we then threshold the ridge-strength values at each chosen percentile and record all connected components found at each percentile threshold. We use the absolute value of the ridge strength because we have found that the sign is not perfectly reliable. Using the absolute value will retain "negative ridges," but we can get rid of these negative ridges later. By recording the percentile-thresholded connected components at each scale, we produce an "inventory" of pixel regions. On a small set of images tried so far, we have found that a reasonable fraction of regions in this inventory correspond to spiral arm segments. We have also found that arm segments are usually represented in this inventory, but tests on much larger image sets will be needed in order to confirm the generality of this finding. If needed, future work could also pursue improvements to inventory coverage by investigating inventory-generation procedures more sophisticated than thresholding.

With a sufficient range and granularity of scale values, spiral arms of all widths will be represented in this inventory. We have also found that with a sufficient range and granularity of the percentile values, we will capture various levels of the spiral arm hierarchy (e.g., a contiguous group of spiral arms contains individual spiral arms, which may contain different parts of a fork, with each arm segment in the fork potentially consisting of "mini-segments" with slight brightness dips in between, which themselves may have mini-segments or a smaller internal contour of stronger brightness). We do not know in advance which scales yield spiral arms (especially since we want our methods to be robust to scale differences), nor do we know in advance which percentile values correspond to which levels in the spiral-structure hierarchy, and the needed scale and percentile values often vary by galaxy. However, if we choose the right pixel regions from this inventory, then it is possible to accommodate all potential spiral arm widths (and lengths) in a spiral arm-segment structure description (to the extent that these widths and lengths are included in the available inventory). Below, we describe our current method for doing so, but this method should be considered a basis for further development and experimentation, not a final procedure.

After generating the inventory, we can filter it to remove regions that obviously do not correspond to spiral arms, as determined by several criteria. This reduces interference from unwanted regions, and reduces the amount of computation needed later (e.g., for least-squares fits). In our experiments so far we use three filtering criteria. First, the regions cannot cover a very large or very small proportion of the image; regions that are too small do not have meaningful structure, and regions that are too large cover galaxy-sized features (e.g., the entire galaxy disk, or entire set of arms) rather than arm-sized features. These minimum and maximum image proportions can be made reliable if an image standardization step is applied (potentially the same one as Section 3.3, except that image resizing is not necessary) so that galaxies occupy roughly consistent proportions of the image. One may also consider making the minimum size an absolute size (in pixels) rather than a proportion. Second, a sufficient proportion of region pixels must have a positive ridge strength (a nonzero proportion cutoff is likely best because the ridge-strength sign can be noisy). Third, when translating the region into polar coordinates, elliptical contours of the covariance matrix should be sufficiently elongated (this filters out bulge and star regions, as well as many regions that are a gross mismatch for a logarithmic spiral).

After this filtering, a logarithmic spiral arc can be fit to each region, in the manner described in Section 3.7.1. We then sort the region inventory by fit quality (best first); we currently use the root mean squared error for this purpose. This gives priority to regions with the best fit. We have also found that the sort operation favors small clusters (to a lesser extent). Next, we can initialize an image that assigns pixels to arm segments, with each pixel initially assigned to the background (i.e., no arm segment). The regions from our inventory can then be considered in sorted order. For each region, we examine all corresponding pixels in the assignment image; if all are background, then we assign all of the corresponding pixels to a new arm segment. So far, we have found that sorting gives us the likely-best arm-segment region occupying that space, so the main remaining concerns are avoiding placements of bad regions, and avoiding interference with legitimate regions that could appear later in the ordering and overlap slightly with previously placed regions. The earlier filtering step removes many bad regions. The remaining bad regions appear later in the region ordering, at which point it is very likely that at least some of the bad region's pixels have already been assigned to an existing arm segment (in which case we are more selective about using the region). Interference with legitimate clusters (appearing later in the ordering) is unlikely because the sorting favors small clusters over large ones (under similar correspondence to a logarithmic spiral pattern), and small clusters represent inner contours of arms. In our experiments so far we have found that overlaps occur at arm "fringes" that are not included in the inner contours, and so the inner contours are unlikely to overlap with each other. However, one could explore adding requirements that a region would need to satisfy in order to create a new arm segment in the assignment image.

If the next region from our sorted inventory overlaps exactly one existing arm segment, the overlap could be because (1) the new region represents a larger contour around the existing arm segment, (2) the new region does not correspond to a spiral arm, or (3) the overlap indicates spatial proximity of two spiral arms at their outer contour levels. In our development so far, we do not make use of the third possibility (it is treated in the same way as the second scenario). Experiments to date have indicated that handling this third scenario is a lower priority. At sufficient granularity of the scale and ridge-strength percentile values, regions are produced at different contours of the same spiral arm, and smaller regions (i.e. inner contours, where between-arm overlap is much less likely) tend to be used first due to the sorting criterion, so region overlaps appear unlikely to cause arm segment omissions. However, further developments could consider adding an arm segment corresponding to the non-overlapping part of the new region, under conditions that would need to be determined (if the conditions are not met, the region would be discarded, or added to supplemental information about the existing arm segment). We then must decide if situation (1) has occurred (in which case we expand the existing arm segment to include the new region) or if situation (2) has occurred (in which case we currently discard the new region, but further development could use this region to calculate supplemental information about the existing arm segment). We currently make this decision based on degree of overlap and change in fit quality, but the optimal details of these decision criteria (and possibly other criteria yet to be considered) still need to be investigated. Our code allows different variations of these criteria to be defined and used interchangeably (for performance comparison).

If the next region from our sorted inventory overlaps two or more existing arm segments, we must decide if all of these arm segments plausibly belong to the same spiral arm. If they do, we combine all existing arm segments overlapping the new region (along with the new region itself) into a new spiral arm segment. We currently make these merge decisions based on the relative fit quality of the original and combined clusters; the exact nature of this decision (or additions or alternatives) still needs to be explored. Again, our code allows strategies to be defined and used interchangeably, facilitating experiments. Regardless of the details of this merge-decision strategy, however, we note that by having larger clusters that overlap multiple smaller clusters (we have found that separate arm segments plausibly belonging to the same spiral arm often appear as a unified arm at larger scales and/or a different ridge-strength contour), we have a powerful way of finding potential merges, including long-distance merges, since the larger cluster fills in gaps between the smaller clusters it overlaps. By filling these gaps, we do not need to interpolate across large cluster gaps, which was the main reason we needed to avoid long-distance cluster merges in Chapter 3.

Instead of only deciding whether to combine *all* arm segments overlapping a later region from the inventory, other possibilities could be considered. For example, a “cluster buster” strategy could divide the new region according to the nearest (or best-fitting) existing region for each pixel, decide which new pixels to keep for each existing region, and then decide on a pairwise basis which existing regions should be merged. In any case, the arm-segment assignment process terminates when the inventory has been exhausted.

Although many details of this plan still need to be determined, we believe that this method is a promising new direction in spiral arm segment extraction. It not only eliminates the assumption that arms within a galaxy have similar widths,[5] but also uses scale variation to its advantage when building and merging clusters. For example, by establishing smaller, high-confidence arm segments first, we can check that larger regions are reasonable (despite a potentially higher fit error that comes with their size and length) by making sure that they agree with previously-established, higher-confidence smaller clusters. Additionally, ridge detections at larger scales can provide the information needed to reliably merge smaller clusters across larger gaps. Another benefit to this method is that it concentrates its computation on logarithmic spiral correspondence rather than also on pixel-level cluster growth. Furthermore, precision/time tradeoffs can be controlled by adjusting the granularity of the scale and percentile values, and one can also adjust the computation-time requirement according to the range of arm scales one needs (or wants) to consider. This method also searches for the best potential arm segment regions in advance, before committing to any assignments of pixels to arm segments.[6] Thus, even though this approach requires further development in order to be usable, and even though this method is not "tried and true" like the one in Chapter 3 (incurring a risk of unexpected problems arising during further development), it is likely worth further exploration.

---

[5] The orientation field uses multiscale information, but the new method described in this section considers more scales, and incorporates logarithmic spiral fit information (rather than pixel-level variance and orientation strength) when combining different scales.

[6] This is currently done using a greedy algorithm, but other types of algorithms could also be explored.

7.3 Brightness Profile Fitting

The list of arm segments (and pixel clusters) produced by our method enables a wealth of information and insights about spiral galaxy structure. However, it would be even better to also obtain automatic fits to the brightness profiles of spiral galaxies, in the same way as is available for the (simpler and less varied) structure of galaxy bulges and disks. Detailed modeling of spiral galaxy brightness is already possible via GALFIT [50, 51], but choosing the appropriate components (along with good initial parameters for these components) is currently an arduous manual task for spiral galaxies. For some science goals, human insight and judgment may be an integral part of modeling and fitting such galaxies, but in other cases, it may be paramount to have impartial and/or large-scale spiral galaxy brightness-pattern information.

Our list-of-arcs output (and the corresponding pixel clusters) could be extremely helpful in facilitating automated fitting of spiral galaxy brightness patterns. Using our output, one could address three main challenges facing automated GALFIT model fitting: interactions between the model parameters, the need to choose model components, and the need to provide an initial guess for the parameters of each component. Interactions between model parameters are problematic where inaccuracies or changes in some parameters disrupt fits of other parameters. Our output could allow increased independence between parameters of different model components. Since the approximate backbone of the spiral arms is known, the arm models would start close enough to the image arms to avoid interfering with fits to other arms. It would also be possible to reduce the influence of the spiral arms on elliptical components such as the galaxy disk; one could fit arm brightness models to the corresponding pixel clusters and subtract this brightness before fitting the disk, or even just down-weight arm-cluster regions during the disk fit. Similarly, our list-of-arcs output could greatly facilitate selection of the model components because the set of arms would be known to a substantial extent. GALFIT also requires good guesses for component parameters. Our output directly provides structural parameters for the spiral arms (these parameters may need to be transformed to suit GALFIT's model conventions), and the corresponding pixel regions can facilitate fitting of the arm segment's brightness parameters.

A system producing GALFIT fits (or initial conditions) from our output could begin by fitting the elliptical (axisymmetric) components, such as the bulge and disk. Elliptical-component fitting capabilities are already established; for example, GALFIT itself could likely be used to automatically fit these components. The bar could likely also be handled in this manner; one would probably be able to determine bar presence by attempting to fit a bar (using a centered elliptical component) and then applying classic model-selection criteria (e.g., the Akaike information criterion). Next, the spiral arms would be added. One could optionally begin fitting the arms by looking for substantial symmetric arm structure in our output; if it exists, the subset of symmetric arms could be fit together, and either used as is or temporarily set aside and compared with the more general individual-arm model.[7] After (possibly) fitting the symmetric arms, the remaining arms from our output (or all of them, if comparing with the symmetric model instead of directly accepting it) would then be introduced, likely one by one and ordered by brightness (using input FITS image brightness values within the pixel cluster from our output) or by the fit quality reported in our output. Since the approximate position and shape of the arm would already be known, along with the pixels most likely to belong to the arm, the fit to the spiral arms' brightness model (including fits potentially made using GALFIT itself) may be greatly facilitated. After fitting each spiral arm (and perhaps refitting previous components in order to make slight updates to their fits) the arm could be accepted if it has sufficient brightness-fit quality (within the arm-cluster region), adds at least a minimum amount of brightness to the overall model, satisfies a model-selection criterion (like the Akaike information criterion), or meets some other test. After fitting and testing all arms in our list-of-arcs output, the resulting

---

[7] It is not necessary to explicitly fit symmetric arms, and the symmetric model is not always appropriate, but it may be informative to look for symmetry and test a symmetric fit.

galaxy brightness model could be used as is, or given to GALFIT as initial conditions.

Alternate means of using our output are, of course, possible. For example, our output likely provides the information needed for a good initialization of an expectation-maximization algorithm [17], which would iteratively alternate between determining the extent to which each pixel is explained by each galaxy component (disk, bulge, bar, arm, etc.), and using the result to perform a weighted fit of each component. Bayesian methods for galaxy fitting (e.g., [52]) may also be promising, either as an intermediate step between our output and GALFIT, or as an alternative to GALFIT.

When building a system for producing GALFIT models (or initial conditions) from our output, some suggestions and insights about manual fitting would likely also apply to automated fitting. A list of such suggestions and insights can be found in the GALFIT User's Manual.[8] These suggestions include: starting with a simple model and gradually adding more components, avoiding most hard constraints on model parameters, and saving computation time by performing an initial fit on a reduced-resolution image.

Once automatically-fitted spiral galaxy brightness models are available, an important application would be to reconcile these observations with theories of spiral galaxy evolution. It would be interesting to fit models that can both explain image brightness and incorporate theories about spiral galaxy formation and evolution. Dynamical grammars [45] likely have the power to do so; these grammars can (among other things) include parameters for the involved components, and can model discrete and continuous changes involving these components and their parameters. As mentioned in Section 2.4, automatic fits have been demonstrated on a few galaxies for a symmetric two-arm galaxy model derived from dynamical grammar rules [70]. It could be very interesting to develop dynamical grammar inference methods that can accommodate general spiral galaxy structure, and that can perform a

---

[8] This document is available at `http://users.obs.carnegiescience.edu/peng/work/galfit/README.pdf`

fit using the grammar rules themselves (producing a reaction path or set of reaction paths leading to the observed galaxy image). Such a procedure could benefit from the information given by an automatic fit to GALFIT (or similar) models of spiral galaxy structure and brightness, and would also benefit (directly or indirectly) from our existing list-of-arcs output.

In all, the spiral arm structure description produced by our method already expands the information available about spiral galaxy structure, especially quantitatively and for large image sets. The information newly available from our method also facilitates important future developments in spiral galaxy structure extraction. We hope our method helps advance astronomical knowledge by enabling large-scale sky surveys to provide the quantitative data needed to address scientific questions about spiral galaxy structure.

Bibliography

[1] B. Aryal. Winding sense of galaxies around the local supercluster. *Research in Astronomy and Astrophysics*, 11(3):293, 2011.

[2] K. Au. *Inferring galaxy morphology through texture analysis.* PhD thesis, Carnegie Mellon University, 2006.

[3] K. Au, C. Genovese, and A. Connolly. Inferring galaxy morphology through texture analysis. Technical Report 843, Carnegie Mellon University, 2006.

[4] N. M. Ball and R. J. Brunner. Data mining and machine learning in astronomy. *International Journal of Modern Physics D*, 19(07):1049–1106, 2010.

[5] S. V. Beckwith, M. Stiavelli, A. M. Koekemoer, J. A. Caldwell, H. C. Ferguson, R. Hook, R. A. Lucas, L. E. Bergeron, M. Corbin, S. Jogee, et al. The Hubble ultra deep field. *The Astronomical Journal*, 132(5):1729, 2006.

[6] J. C. Berrier, B. L. Davis, D. Kennefick, J. D. Kennefick, M. S. Seigar, R. S. Barrows, M. Hartley, D. Shields, M. C. Bentz, and C. H. Lacy. Further Evidence for a Supermassive Black Hole Mass-Pitch Angle Relation. *The Astrophysical Journal*, 769(2):132, 2013.

[7] E. Bertin and S. Arnouts. SExtractor: Software for source extraction. *Astronomy and Astrophysics Supplement Series*, 117:393–404, 1996.

[8] T. Borchkhadze and N. Kogoshvili. The anisotropy of spiral galaxy orientation. *Astronomy and Astrophysics*, 53:431–433, 1976.

[9] C. Cardamone, K. Schawinski, M. Sarzi, S. P. Bamford, N. Bennert, C. Urry, C. Lintott, W. C. Keel, J. Parejko, R. C. Nichol, et al. Galaxy Zoo Green Peas: discovery of a class of compact extremely star-forming galaxies? *Monthly Notices of the Royal Astronomical Society*, 399(3):1191–1205, 2009.

[10] C.-G. Danver. A Morphological Investigation of Some Near Galaxies with Regard to the Lengths and the Form of their Arms, their Inclinations and their Symmetry Properties. *Annals of the Observatory of Lund*, 10:3–193, 1942.

[11] B. L. Davis, J. C. Berrier, D. W. Shields, J. Kennefick, D. Kennefick, M. S. Seigar, C. H. Lacy, and I. Puerari. Measurement of Galactic Logarithmic Spiral Arm Pitch Angle Using Two-dimensional Fast Fourier Transform Decomposition. *The Astrophysical Journal Supplement Series*, 199(2):33, 2012.

[12] D. R. Davis and W. B. Hayes. Automated quantitative description of spiral galaxy arm-segment structure. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1138–1145. IEEE, 2012.

[13] D. R. Davis and W. B. Hayes. SpArcFiRe: Scalable Automated Detection of Spiral Galaxy Arm Segments. *The Astronomical Journal*, 2014. submitted.

[14] M. E. de Souza. A new model without dark matter for the rotation of spiral galaxies: the connections among shape, kinematics and evolution. *Frontiers in Science*, 3(2):71–80, 2013.

[15] R. De Souza, D. A. Gadotti, and S. Dos Anjos. BUDDA: a new two-dimensional bulge/disk decomposition code for detailed structural analysis of galaxies. *The Astrophysical Journal Supplement Series*, 153(2):411, 2004.

[16] G. De Vaucouleurs. Revised Classification of 1500 Bright Galaxies. *The Astrophysical Journal Supplement Series*, 8:31, 1963.

[17] A. Dempster, N. Laird, and D. Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.

[18] R. O. Duda and P. E. Hart. Use of the Hough transformation to detect lines and curves in pictures. *Communications of the ACM*, 15(1):11–15, 1972.

[19] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(9):1627–1645, 2010.

[20] P. F. Felzenszwalb and D. P. Huttenlocher. Pictorial structures for object recognition. *International Journal of Computer Vision*, 61(1):55–79, 2005.

[21] A. Fridman and S. Poltorak. Using an assumption about the monotony of spiral arms to determine the orientation angles of galaxies. *Monthly Notices of the Royal Astronomical Society*, 403(3):1625–1632, 2010.

[22] J. Gallagher, L. Sparke, and L. Matthews. Disk Galaxies. In P. Murdin, editor, *Encyclopedia of Astronomy and Astrophysics*. Nature Publishing Group, Oct. 2002.

[23] C. Garcia-Gomez and E. Athanassoula. Analysis of the distribution of HII regions in external galaxies. I, Position and inclination angles. *Astronomy & astrophysics. Supplement Series*, 89(1):159–184, 1991.

[24] B. Häußler, S. P. Bamford, M. Vika, A. L. Rojas, M. Barden, L. S. Kelvin, M. Alpaslan, A. S. Robotham, S. P. Driver, I. Baldry, et al. MegaMorph–multiwavelength measurement of galaxy structure: complete Sérsic profile information from modern surveys. *Monthly Notices of the Royal Astronomical Society*, 430(1):330–369, 2013.

[25] L. C. Ho, Z.-Y. Li, A. J. Barth, M. S. Seigar, and C. Y. Peng. The Carnegie-Irvine Galaxy Survey. I. Overview and Atlas of Optical Images. *The Astrophysical Journal Supplement Series*, 197(2):21, 2011.

[26] T. Hu, Q.-H. Peng, and H.-S. Zong. Disk Thicknesses and Some Parameters of 108 Non-Edge-On Spiral Galaxies. *Chinese Journal of Astronomy and Astrophysics*, 7(4):579, 2007.

[27] T. Hu, Z. Shao, and Q. Peng. New Estimates of the Inclination, Position Angle, Pitch Angle, and Scale Height of the Whirlpool Galaxy. *The Astrophysical Journal Letters*, 762(2):L27, 2013.

[28] T. Huang, G. Yang, and G. Tang. A fast two-dimensional median filtering algorithm. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 27(1):13–18, 1979.

[29] E. P. Hubble. *The realm of the nebulae*. Yale University Press, 1936.

[30] G. D. Illingworth, P. Oesch, R. Bouwens, and X. Team. The XDF: Hubble's eXtreme Deep Field. In *American Astronomical Society Meeting Abstracts*, volume 221, 2013.

[31] M. Iye and H. Sugai. A catalog of spin orientation of southern galaxies. *The Astrophysical Journal*, 374:112–116, 1991.

[32] A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: a review. *ACM computing surveys (CSUR)*, 31(3):264–323, 1999.

[33] D. K. Kondepudi and D. J. Durand. Chiral asymmetry in spiral galaxies? *Chirality*, 13(7):351–356, 2001.

[34] K. Land, A. Slosar, C. Lintott, D. Andreescu, S. Bamford, P. Murray, R. Nichol, M. J. Raddick, K. Schawinski, A. Szalay, et al. Galaxy Zoo: the large-scale spin statistics of spiral galaxies in the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society*, 388(4):1686–1692, 2008.

[35] L. Levi. Unsharp masking and related image enhancement techniques. *Computer Graphics and Image Processing*, 3(2):163–177, 1974.

[36] C. Lin and F. H. Shu. On the Spiral Structure of Disk Galaxies. *The Astrophysical Journal*, 140:646, 1964.

[37] T. Lindeberg. Edge detection and ridge detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):117–156, 1998.

[38] C. Lintott, K. Schawinski, S. Bamford, A. Slosar, K. Land, D. Thomas, E. Edmondson, K. Masters, R. C. Nichol, M. J. Raddick, et al. Galaxy Zoo 1: data release of morphological classifications for nearly 900 000 galaxies. *Monthly Notices of the Royal Astronomical Society*, 410(1):166–178, 2011.

[39] C. J. Lintott, K. Schawinski, W. Keel, H. Van Arkel, N. Bennert, E. Edmondson, D. Thomas, D. J. Smith, P. D. Herbert, M. J. Jarvis, et al. Galaxy Zoo:'Hanny's Voorwerp', a quasar light echo? *Monthly Notices of the Royal Astronomical Society*, 399(1):129–140, 2009.

[40] C. J. Lintott, K. Schawinski, A. Slosar, K. Land, S. Bamford, D. Thomas, M. J. Raddick, R. C. Nichol, A. Szalay, D. Andreescu, et al. Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society*, 389(3):1179–1189, 2008.

[41] M. J. Longo. Does the Universe Have a Handedness? *arXiv preprint arXiv:0812.3437*, 2008.

[42] M. J. Longo. Detection of a Dipole in the Handedness of Spiral Galaxies with Redshifts z$\sim$0.04. *Physics Letters B*, 699(4):224–229, 2011.

[43] R. Lupton, M. R. Blanton, G. Fekete, D. W. Hogg, W. OMullane, A. Szalay, and N. Wherry. Preparing Red-Green-Blue Images from CCD Data. *Publications of the Astronomical Society of the Pacific*, 116(816):133–137, 2004.

[44] J. Ma. A method of obtaining the pitch angle of spiral arms and the inclination of galactic discs. *Chinese Journal of Astronomy and Astrophysics*, 1(5):395, 2001.

[45] E. Mjolsness and G. Yosiphon. Stochastic process semantics for dynamical grammars. *Annals of Mathematics and Artificial Intelligence*, 47(3-4):329–395, 2006.

[46] P. B. Nair and R. G. Abraham. A catalog of detailed visual morphological classifications for 14,034 galaxies in the sloan digital sky survey. *The Astrophysical Journal Supplement Series*, 186(2):427, 2010.

[47] I. I. Pasha. Sense of Rotation in 109 Spirals and the Leading Arms in the Interacting Galaxies NGC3786 and NGC5426. *Soviet Astronomy Letters*, 11:1–4, Jan. 1985.

[48] I. Patrikeev, A. Fletcher, R. Stepanov, R. Beck, E. Berkhuijsen, P. Frick, and C. Horellou. Analysis of spiral arms using anisotropic wavelets: gas, dust and magnetic fields in M 51. *Astronomy and Astrophysics*, 458:441–452, 2006.

[49] C. Y. Peng. GALFIT User's Manual. `http://users.obs.carnegiescience.edu/peng/work/galfit/README.pdf`.

[50] C. Y. Peng, L. C. Ho, C. D. Impey, and H.-W. Rix. Detailed structural decomposition of galaxy images. *The Astronomical Journal*, 124(1):266, 2002.

[51] C. Y. Peng, L. C. Ho, C. D. Impey, and H.-W. Rix. Detailed decomposition of galaxy images. II. Beyond axisymmetric models. *The Astronomical Journal*, 139(6):2097, 2010.

[52] B. Perret, V. Mazet, C. Collet, and É. Slezak. Hierarchical multispectral galaxy decomposition using a MCMC algorithm with multiple temperature simulated annealing. *Pattern Recognition*, 44(6):1328–1342, 2011.

[53] I. Reid, C. Brewer, R. Brucato, W. McKinley, A. Maury, D. Mendenhall, J. Mould, J. Mueller, G. Neugebauer, J. Phinney, et al. The second palomar sky survey. *Publications of the Astronomical Society of the Pacific*, 103:661–674, 1991.

[54] B. Ripley and A. Sutherland. Finding spiral structures in images of galaxies. *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences*, 332(1627):477–485, 1990.

[55] M. S. Roberts and M. P. Haynes. Physical parameters along the Hubble sequence. *Annual Review of Astronomy and Astrophysics*, 32:115–152, 1994.

[56] M. Saraiva Schroeder, M. Pastoriza, S. Kepler, and I. Puerari. The distribution of light in the spiral galaxy NGC 7412. *Astronomy and Astrophysics Supplement Series*, 108:41–54, 1994.

[57] S. Savchenko. On estimation of the pitch angle for spiral arms of galaxies. *Astrophysical Bulletin*, 67(3):310–317, 2012.

[58] S. Savchenko and V. Reshetnikov. Pitch angles of distant spiral galaxies. *Astronomy letters*, 37(12):817–825, 2011.

[59] S. Savchenko and V. Reshetnikov. Pitch angle variations in spiral galaxies. *Monthly Notices of the Royal Astronomical Society*, 436(2):1074–1083, 2013.

[60] M. S. Seigar, J. S. Bullock, A. J. Barth, and L. C. Ho. Constraining dark matter halo profiles and galaxy formation models using spiral arm morphology. I. Method outline. *The Astrophysical Journal*, 645(2):1012, 2006.

[61] L. Shamir. Ganalyzer: A tool for automatic galaxy image analysis. *The Astrophysical Journal*, 736(2):141, 2011.

[62] L. Shamir. Automatic detection of peculiar galaxies in large datasets of galaxy images. *Journal of Computational Science*, 3(3):181–189, 2012.

[63] L. Shamir. Handedness asymmetry of spiral galaxies with z<0.3 shows cosmic parity violation and a dipole axis. *Physics Letters B*, 715:25–29, 2012.

[64] L. Simard. GIM2D: an IRAF package for the Quantitative Morphology Analysis of Distant Galaxies. In *Astronomical Data Analysis Software and Systems VII*, volume 145, page 108, 1998.

[65] L. A. Thompson. Spiral galaxy orientation in the Virgo Cluster. *Publ. Astron. Soc. Pac., v. 85, no. 507, p. 528*, 85(507), 1973.

[66] S. Van den Bergh. *Galaxy morphology and classification.* Cambridge University Press, 1998.

[67] K. W. Willett, C. J. Lintott, S. P. Bamford, K. L. Masters, B. D. Simmons, K. R. Casteels, E. M. Edmondson, L. F. Fortson, S. Kaviraj, W. C. Keel, et al. Galaxy Zoo 2: detailed morphological classifications for 304 122 galaxies from the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society*, 435(4):2835–2860, 2013.

[68] T. Yamagata, M. Hamabe, and M. Iye. Angular momentum orientation of spiral galaxies. *Tokyo, University, Tokyo Astronomical Observatory, Annals, Second Series*, 18(3):164–174, 1981.

[69] D. G. York, J. Adelman, J. E. Anderson Jr, S. F. Anderson, J. Annis, N. A. Bahcall, J. Bakken, R. Barkhouser, S. Bastian, E. Berman, et al. The sloan digital sky survey: Technical summary. *The Astronomical Journal*, 120(3):1579, 2000.

[70] G. Yosiphon. *Stochastic Parameterized Grammars: Formalization, Inference and Modeling Applications.* PhD thesis, University of California, Irvine, 2009.

[71] H. Yuen, J. Princen, J. Illingworth, and J. Kittler. Comparative study of Hough transform methods for circle finding. *Image and Vision Computing*, 8(1):71–77, 1990.

[72] S.-C. Zhu and D. Mumford. A Stochastic Grammar of Images. *Foundations and Trends in Computer Graphics and Vision*, 2(4):259–362, 2006.