

The Pennsylvania State University  
The Graduate School

AESTHETIC AND AFFECTIVE APPLICATIONS OF VISUAL  
FEATURES AND PHYSICAL PROPERTY ANALYSIS OF PLANT  
STOMATA THROUGH ACTIVE SURFACES

A Dissertation in  
Information Sciences and Technology  
by  
Barış Kandemir

© 2019 Barış Kandemir

Submitted in Partial Fulfillment  
of the Requirements  
for the Degree of

Doctor of Philosophy

May 2019

The dissertation Barış Kandemir was reviewed and approved\* by the following:

James Z. Wang  
Professor, College of Information Sciences and Technology  
Dissertation Advisor, Co-Chair of Committee

Jia Li  
Professor, Department of Statistics  
Co-Chair of Committee

Zihan Zhou  
Assistant Professor, College of Information Sciences and Technology

Reginald B. Adams  
Professor, Department of Psychology

Charles T. Anderson  
Associate Professor, Department of Biology

Mary Beth Rosson  
Professor, College of Information Sciences and Technology  
Graduate Programs Head

\*Signatures are on file in the Graduate School.

# Abstract

The rapid advancements in mobile and standalone imaging technologies have made it possible to obtain visual data faster and more easily. This has led to an explosion in the amount of online (i.e. social media, etc.) and offline (i.e. scientific data in biology and astronomy) visual data. Hence, a need to manage data in an acceptable time frame via reasonable methods emerged. Users of online media require rapid content analysis and creation, which requires personalization and automation. In parallel, time and accuracy are both important constraints to today's scientists as they go through substantial number of images to test their hypotheses. People have sought help from computerized systems in order to catch up with the speed of the data. In content creation, computational aesthetic and affective studies strive to create systems that help understand and predict people's collective response to online visual content, or create better content that address more people. In similar vein, manual analysis of biological structures and their physical properties has become a bottleneck for biomedical research with the increase in the amount of scientific visual data. One such example is stomatal research in plant biology. Scientists analyze subcellular structures and physical properties of stomatal guard cells of plants which dynamically change shape to control photosynthesis and water transport. Fluorescence microscopy is one of the most frequent methods utilized to analyze the stomatal complex in 3D with its own challenges.

In this dissertation, we try to answer requirements for automation of both visual content creation and biomedical visual data analysis. For visual content analysis and creation, we consider visual features that are commonly used for computational aesthetics and affective studies. For biomedical analysis, we consider automation of detection of tubular subcellular structures and 3D cell surface analysis for stomatal research. In these lines, we investigate aesthetic and affective differences in response to these features from different demographic groups and attempt to improve these features. A study has been conducted to understand commonalities and differences across latent demographic groups in crowdsourcing participant base. The differences in aesthetic and affective response to visual features in images, such as color and simplicity-complexity, and emotional qualities evoked by them, such as valence, arousal and dominance was conducted. In order to improve aforementioned visual features, a small scaled study was conducted

about compositional balance. The study reveals that visual saliency, which is commonly used for visual balance representation, does not work very well. Human positions in images, perspective points can be utilized for a better visual balance representation.

In case of stomatal research, scientists focus on different subcellular structures and general 3D guard cell geometry for different states of stoma complex. Microtubules (MTs) are subcellular structures that play important role in cellulose deposition. However, high number and density of MTs, contrast loss in images due to their position relative to the imaging plane, and their crossovers render the manual analysis challenging for biologists. This dissertation work proposes a method to automatically extract MT locations from 2D projections of confocal micrographs. Our method uses a tensor voting framework to guide stretching open active curves (SOACs) and improve curve continuity when there is a loss of contrast. We utilize 2D maximum intensity projections of spinning disk confocal microscopy image stacks of fluorescent protein (FP)-labeled MTs collected from guard cells in *Arabidopsis thaliana* seedlings to validate the benefits of adding a tensor voting framework. A survey was created to compare our results to the baseline, where participants from a university plant biology department chose between randomized results. Results demonstrated that our method performs nearly consistently better than the baseline. Automation can also be employed for general 3D guard cell geometry analysis as 3D models provide valuable insight about their functioning. We present a semi-automated 3D surface segmentation program that provides a 3D model of stomatal complex through segmentation by polar active contours. As active contours can handle topological changes and small sized gaps, the connection regions to surrounding pavement cells are smoothly closed delivering a completely isolated stoma structure in 3D. This tool is shown to obtain 3D stoma model with little human intervention.

In summary, our results from these studies show that visual content creation can be accelerated through personalization and better image analysis. Findings lead to a better understanding of the crowdsourcing participants, hence better survey design. In parallel, image segmentation techniques can be leveraged for fast and reliable analysis of biomedical data. The proposed segmentation method can be applied to a wide range of biological images with tubular structures. The obtained 3D model can be put to use for geometrical analysis of stomata at different states. Material properties such as elasticity can be better analyzed, as the inner and outer walls are automatically distinguished.



# Table of Contents

<b>List of Figures</b>	<b>ix</b>
<b>List of Tables</b>	<b>xiv</b>
<b>Acknowledgments</b>	<b>xv</b>
<b>Chapter 1</b>	
<b>Introduction</b>	<b>1</b>
1.1 Viewer Perspective: Compositional Aesthetics . . . . .	3
1.2 Viewer Perspective: Emotions . . . . .	5
1.3 Plant Biology Perspective . . . . .	6
1.3.1 Automatic Segmentation of Subcellular Structures . . . . .	6
1.3.2 3D Surface Extraction of the Stomatal Complex . . . . .	9
1.4 Problem Statement . . . . .	10
1.4.1 Research Questions . . . . .	10
1.4.2 Contributions . . . . .	11
1.5 Structure of Dissertation . . . . .	12
<b>Chapter 2</b>	
<b>Background: Computational Aesthetics and Affective Computing</b>	<b>13</b>
2.1 Introduction . . . . .	13
2.2 Aesthetics . . . . .	13
2.3 Balance and Aesthetics in Multimedia . . . . .	19
2.3.1 Symmetry . . . . .	20
2.3.2 Texture . . . . .	20
2.3.3 Shape . . . . .	21

2.3.4	Color and Value . . . . .	22
2.3.5	Depth . . . . .	22
2.3.6	Saliency . . . . .	22
2.4	Emotions and Affect . . . . .	25
2.5	Emotional and Preferential Differences Across Different Demographics in Crowdsourcing . . . . .	28
2.5.1	The Utilization of Crowdsourcing for Labeled Data . . . . .	28
2.5.2	Amazon Mechanical Turk (AMT) . . . . .	29
2.5.3	Studies on Differences Across Various Demographics . . . . .	30
2.5.4	Personalization . . . . .	30
2.6	Summary . . . . .	31

### Chapter 3

	<b>Background: Tubular Structure Analysis in Bioimaging</b>	<b>32</b>
3.1	Introduction . . . . .	32
3.2	Stomata Research in Plant Biology . . . . .	32
3.2.1	Microtubule Research in Plant Biology . . . . .	33
3.2.2	3D Surface Modeling . . . . .	33
3.3	Morphological Operations . . . . .	34
3.4	Image Denoising in Subcellular Structures . . . . .	36
3.4.1	Low Rank Image Representation . . . . .	37
3.4.2	Diffusion Filters . . . . .	38
3.5	Tubular Structure Segmentation . . . . .	42
3.5.1	Filter/Model Based Methods . . . . .	42
3.5.2	Energy Minimization . . . . .	47
3.5.3	Learning Based Methods . . . . .	50
3.6	Perceptual Grouping-Tensor Voting . . . . .	51
3.7	Summary . . . . .	52

### Chapter 4

	<b>Beyond Saliency: Assessing Visual Balance with High-level Cues</b>	<b>53</b>
4.1	Introduction . . . . .	53
4.2	The Method . . . . .	55
4.2.1	Data Collection . . . . .	56
4.2.2	High- and Low-Level Features for Balance . . . . .	58
4.2.3	Representation . . . . .	61
4.3	Evaluation . . . . .	62
4.4	Discussion . . . . .	66
4.5	Summary . . . . .	68

<b>Chapter 5</b>	
<b>Differences and Biases in Affect Evoked by Visual Features in Latent Crowdsourcing Demographic Groups</b>	<b>69</b>
5.1 Introduction . . . . .	69
5.2 Method . . . . .	70
5.2.1 Data Collection . . . . .	70
5.2.2 Visual Features . . . . .	74
5.2.3 Data Analysis . . . . .	76
5.3 Findings . . . . .	81
5.3.1 Commonalities Across Groups . . . . .	82
5.3.2 Differences Across Groups . . . . .	83
5.4 Discussion . . . . .	84
5.5 Summary . . . . .	84
<b>Chapter 6</b>	
<b>Microtubule Segmentation and Guard Cell Pair Isolation</b>	<b>85</b>
6.1 Introduction . . . . .	85
6.2 Method . . . . .	85
6.2.1 Frangi Vesselness Measure for 2D . . . . .	86
6.2.2 Preprocessing . . . . .	87
6.2.3 Heuristics Initialization . . . . .	88
6.2.4 Tubular Geometry Aware SOACs (G-SOAX) . . . . .	89
6.2.5 Artifact Removal . . . . .	91
6.3 Evaluation . . . . .	92
6.3.1 The Microtubule Image Dataset . . . . .	92
6.3.2 Survey-based Evaluation . . . . .	93
6.3.3 Results . . . . .	94
6.3.4 Generalizability to Other Tubular Structure Data . . . . .	95
6.4 Discussion . . . . .	97
6.5 Summary . . . . .	99
<b>Chapter 7</b>	
<b>3D Inner and Outer Wall Extraction of Stomata Guard Cell Pair Through Active Surfaces</b>	<b>100</b>
7.1 Introduction . . . . .	100
7.2 Method . . . . .	101
7.2.1 Manual Contrast Enhancement . . . . .	101
7.2.2 Initial Surface Extraction . . . . .	102
7.2.3 Inner and Outer Wall Separation . . . . .	105
7.2.4 Final Surface Smoothing . . . . .	108

7.2.5	Rendering and Manual Postprocessing . . . . .	108
7.3	Evaluation . . . . .	108
7.3.1	Data . . . . .	108
7.3.2	Results . . . . .	109
7.4	Discussion . . . . .	109
7.5	Summary . . . . .	112
<b>Chapter 8</b>		
	<b>Conclusion and Future Work</b>	<b>113</b>
8.1	Summary . . . . .	113
8.2	Areas of Future Work . . . . .	114
8.2.1	Improvements to Balance Study . . . . .	114
8.2.2	Integration of Balance into Composition Retrieval . . . . .	115
8.2.3	Personalization and Crowdsourcing Survey Design . . . . .	116
8.2.4	Stomatal Research . . . . .	117
8.3	Anticipated Outcomes and Implications . . . . .	119
<b>Appendix A</b>		
	<b>Secondary Research For Stomatal Analysis</b>	<b>120</b>
A.1	Rosette Area Segmentation . . . . .	120
A.2	Stoma Geometrical Properties Analysis . . . . .	121
	<b>Bibliography</b>	<b>125</b>

# List of Figures

1.1	Balance in composition is an important aesthetic concept. It can be employed in different ways to have different visual impacts on viewers. The first image shows an example of symmetric balance, whereas the second image is imbalanced. The third image is imbalanced due to the tree trunk. The fourth image is balanced, as the fence is leading the eye away from the tower. . . . .	3
1.2	Microfibrils transversely cover the cell, constraining lateral growth [1]. . . . .	7
2.1	The off-center disk in a square frame and the so-called structural net with equilibrium points [2]. . . . .	15
2.2	One of the original images in the study (a) and a modified version of it wherein the gamma ramp was applied (b). The white point represents the center of mass (CoM) [3]. . . . .	17
2.3	The elements that can be used in balance from upper-left to lower-right: value, color, shape, texture, position of object, and eye direction [4]. . . . .	18
2.4	2D circumplex proposed in [5] and 3D emotional space [6]. . . . .	26
3.1	The effects of opening and closing on a 1D intensity profile (3.1a) and on a sample image, respectively (3.1b). For the intensity profile, opening eats away the spurious part, whereas closing fills up the valleys between intensity spikes. This behavior can be observed in the image in (3.1b). The original image is given in the upper-right hand corner. The lower-left corner shows the results of a grayscale opening, where small bright details are lost with no great effect on the dark details. The lower-right corner indicates the results of closing, where small dark details are lost with no great effect on the brighter details [7]. . . . .	34
3.2	Point spread function [8]. . . . .	36

3.3	Coherence enhancing diffusion applied to Van Gogh’s ‘Road with Cypress and Star’ [9]. . . . .	41
3.4	Illustration of the eigenvectors $V_i$ (Column 1) and corresponding eigenvalues $\mu_i$ (Column 2) of the structure tensor in five different situations. The first case indicates almost no contrast in any direction; the eigenvalues are equal and close to 0. The second case shows contrast in the direction of $V_1$ ; the first eigenvalue is much larger than the other two eigenvalues. The third case shows contrast in the direction of $V_1$ and $V_2$ ; the first two eigenvalues are much larger than the third. In the fourth case, there is great contrast in $V_1$ , less contrast in the direction of $V_2$ , and almost no contrast in the direction of $V_3$ . The fifth case indicates contrast in all directions; the eigenvalues are equal but are much larger than 0. . . . .	45
3.5	The osculating circle connecting the voting point and voting location [10]. . . . .	51
3.6	The stick voting field and its propagation [10]. . . . .	52
4.1	Saliency analyses often fail to capture the key elements in attention. Columns: original image, Achanta, GBVS, DRFI, Context aware, imSig, covSal, corSal, UHF, SWD, Murray, FES, MDF. . . . .	54
4.2	Our framework for visual center prediction. . . . .	55
4.3	User interface for online human subject study. The slider bar below the image allows the participant to indicate the visual center. . . .	56
4.4	Sample images with low and high standard deviations for the average slider position. Images with lower standard deviations are shown in the first row, where slider positions overlap and the position is illustrated with a single triangle. The second row shows images with higher standard deviations. The participant ratings are marked using triangles to show the spread of the slider positions.	59
4.5	Eye leading lines are utilized to strike a balanced composition[4]. . .	60
4.6	Representation of different features. Gaussians are fitted within the boxes obtained and the CoM of this image is calculated. From top-left to bottom-right: YOLO human detection, text detection, dominant VP detection, and the corresponding Gaussian envelopes.	61
4.7	MSE values of Model 1, Model 2, Model 3, and Model 4 for selected saliency algorithms. Each line corresponds to a saliency algorithm chosen to represent the categories given in Table2.1. . . . .	64

4.8	Evaluation on visual center prediction. The yellow triangle shows mean visual center position annotated by participants. The green triangle is the CoM of saliency images. The red triangle is the prediction of the proposed model. The top row demonstrates cases where our model improved prediction along with the saliency maps (context aware saliency). The lower row displays cases where the model was worse than saliency or didn't improve prediction. . . . .	65
5.1	Ethnicity, income, education, and age distribution of AMT participants after data cleaning. . . . .	71
5.2	Survey Webpage used for human subject study [11]. SAMs are mapped to slider values so that users can employ the sliders to record the values that reflect their emotions. . . . .	73
5.3	Latent representative demographic groups and their make-up in terms of income, ethnic group, education, gender, and age. . . . .	78
5.4	A sample of a topic whose most dominant words were purple, red, and white. Each image that belongs to this topic is ordered via the vectorial norm of the three features' values. Lower-triangle images demonstrate these features better. . . . .	80
6.1	Improvements benefiting from a heuristic initialization. The second column shows seed points through tangential local maxima as opposed to ridge detection in the first row. . . . .	89
6.2	The principle eigenvector field of voting tensor $\mathcal{T}$ (blue arrows) overlaid on top of the vesselness map, and curl points below the mean curl magnitude overlaid on the original sample in red. Blue dots represent initial curve points extracted through our tangential tracing heuristic. . . . .	90
6.3	Comparison of normal SOAX (first row) versus curl and vesselness-driven G-SOAX (second row) in the same samples. The ellipses point to regions where G-SOAX continued evolving to combine the snakes. . . . .	91
6.4	The artifact removal problem becomes easier as the resulting snakes are converted to polar coordinates. The boundary artifacts and tubules separate into vertical and horizontal structures. The use of a Sobel edge filter applied to the binary image can delineate these structures. The bounding box of connected components provides an idea of the structure orientation. The magenta bounding boxes show horizontal structures in both of the thresholded horizontal and vertical difference images. . . . .	92

6.5	Some results from the online evaluation survey. The samples with boundary artifacts are from the first part of the survey and the cleared results are from the second part. The first row shows the G-SOAX results, while the second row shows the SOAX results. The first three columns are samples where the G-SOAX results were preferred, whereas the last two columns are samples where the SOAX results were preferred. It can be observed in the first three columns that G-SOAX helped to create more regular results free of unnecessary branches. . . . .	94
6.6	Comparison of the Aster data, where black structures are MTs. The first column shows the original images. The second and third columns are tubule tracing results for SOAX run at 100 and 10,000 maximum iterations, respectively. The last column shows the G-SOAX results. The blue circles point to exemplary regions where tubule continuity is better with our algorithm. . . . .	96
7.1	The pipeline of our method. The first step is manual contrast enhancement, which is followed by initial active surface segmentation. The third step is the extraction of the inner and outer walls while closing the gaps due to imaging methods. The final step involves the manual manipulation and rendering of the extracted surface into meshes for visualization. . . . .	101
7.2	Contrast enhancement difference. The borders towards the bottom are indiscernible. . . . .	102
7.3	Initial surface extraction through 3D active surfaces. A cube that is the initial surface evolves according to guard cell shape leading to the initial surface segmentation. The second step improves this crude segmentation. The stoma guard cell pair is shown on XY (top-left), YZ (top-right), XZ planes at the yellow crosshair position.	104
7.4	The low contrast areas cause segmentation errors as the segmentation surface collapse into the guard cell, as seen in first row. In order to get the whole wall system, we run the same segmentation with two times with lower (second row) and higher (first row) $\mu$ values from Eq.(7.2). The lower $\mu$ values causes the detection of pavement cells and noise, while higher $\mu$ presents tighter fit with a probability of inward collapse. . . . .	105
7.5	Top: Polar snake in polar coordinates $(\rho, \Theta)$ . Bottom: Polar snake transformed back to Cartesian coordinates. This process is conducted for each slice in volume of interest (VOI). Note the bulbous region on the inner side of the contour. . . . .	107



7.6	Results on the collected samples. The first four rows demonstrate the process of manual manipulation and smoothing after the surface points are obtained and resulting guard cell pair. Bottom row compares our results to other models that are recently employed in [12, 13]. The actual close stoma XY slice shows some deflations on sides which are captured by our model. . . . .	110
8.1	Using the calculated balance point position, real-time feedback can be given to users [14]. . . . .	116
A.1	The segmentation of leaves by thresholding according to their hue, saturation and value(brightness). . . . .	121
A.2	Segmentation of pore are in 3D through semi-automated 2D snakes. The pore area is segmented using snakes algorithm through an initial curve across slices (C-D). The distribution of pore areas is first cleaned from outliers, then a robust curve is fitted, and minimum is detected (E-F). . . . .	122
A.3	Single cell including pore area is detected in a similar fashion to pore area segmentation, and the associated pore segmentation is subtracted to obtain guard cell (B-C-D). The junction point locations are determined via tangential tracing of intensity values (E-F). Radial tracing of intensity is conducted to detect cell walls, hence to find the medial tangential axes (G-H). . . . .	123

# List of Tables

2.1	State-of-the-art bottom-up saliency methods summarized according to their approaches. . . . .	24
3.1	The structure analysis is based on the eigenvalues of the Hessian matrix when the eigenvalues are ordered as $ \lambda_1  \leq  \lambda_2  \leq  \lambda_3 $ , where a positive sign indicates positiveness. This classification applies to bright objects against a dark background. For dark objects on a bright background, the signs should be reversed. . . . .	44
5.1	The affective and aesthetic features utilized in gender difference analysis. . . . .	74
6.1	Structure analysis based on the eigenvalues of the Hessian matrix, when the eigenvalues are ordered as $ \lambda_1  \leq  \lambda_2 $ . This classification applies to bright tubular objects against a dark background. . . . .	87
6.2	Detailed analysis of the survey in terms of the ratio of the number of samples. Numbers in bold indicate ratios where G-SOAX was preferred over SOAX. SOAX was chosen more in the first phase of the survey. However, our method was chosen with a dominating ratio for the survey's second part, which considers the overall quality of the segmentation. . . . .	94

# Acknowledgments

First and foremost, I would like to take the opportunity to thank my advisor, Prof. Dr. James Z. Wang. This thesis work would not be possible without his continuous guidance, support and constant push for improvement during my years at Penn State. Prof. Wang provided me with the skills to be an independent researcher. I learned from him until the last second of my Ph.D.

I would like to thank Dr. Jia Li, Dr. Charles T. Anderson, and Dr. Reginald B. Adams as my committee members for their guidance and suggestions for my research. Many thanks to our collaborators Hanjoo Kim, Dr. Michelle G. Newman, Dr. Junping Zhang, Dr. Yue Rui, Dr. Hojae Yi, and Dr. Virendra Puri. I would also like to thank my labmates Dr. Xin Lu, Dr. Neela Sawant, Dr. Yu Zhang, Dr. Jianbo Ye, Farshid Farhad, and Mohammad Kamani. Thank you to those undergraduate students involved in my research, Edward Chen, Kun Wang, Frederick Zhang, and Mihir Alve. It has been an honor to work with all of you.

Thank you to Yu Zhang, Lin Yang, Derek Miller, and Mark Wheeler for industry experience during my internship. The opportunity to work on real-world industry problems was inspiring and widened my horizons.

Thanks to the small, but lovely town of State College. I never would have thought that there would be a lot to experience in such a small town. The people and the community will always hold a special place in my heart. Thank you to all the wonderful friends that I have made over the years. There are too many to mention, but I would like to say a special thank you to Kyle M. Williams, Pierre Turié, Mehmet Ünal, Shakil Rabbi, Giuseppe Petracca and Andrew Whalen. I will always cherish the special times we shared together. Finally, I would like to thank my family. To my parents, Şükrü, and Servet, and my brother Umut Kandemir. I had to leave home to complete this part of my life, but it was always with the knowledge that you were with me and supporting me without any doubt. Special thank you to Emel Kuruoğlu for her support and kindness.

This material is based upon work supported by the National Science Foundation under Grant Nos. 1027854, 1110970, and 1616316.

# Dedication

To my mother Servet Kandemir, who didn't have chance to see her son walk her path,

To my father Şükrü Kandemir, for his unconditional support.

# Chapter 1

## Introduction

*Beauty in things exists in the mind which contemplates them.* - David Hume

In accordance with the well-known observation that is Moore's law, computing power has dramatically increased over past two decades [15]. The increase has led to many opportunities, one of which is advancements in imaging technology. These advancements have allowed for faster and easier image acquisition. Mobile devices with relatively high imaging capabilities have led to an explosion in the amount of visual data available online. This phenomenon is referred to as "big data," owing to its velocity, volume, and variety [16]. Another such advancement is the advent of Web 2.0, which has enabled the fast dissemination of multimedia. Many national libraries, art galleries, and museums, among other institutions, have seized this opportunity and digitized their collections and archives<sup>1</sup>. Similarly, the amount of scientific visual data available has increased in fields such as biology and astronomy. With the advent of Web 2.0, the amount of data has become hard to manage. Older quantitative techniques for analyzing data failed to keep up with the quantity and the speed of the data available.

A spirit of collaboration has grown up alongside technological developments such as the Internet. More people with access to the Internet has led to a phenomenon called crowdsourcing. Researchers and certain companies have started to outsource tasks that require human intelligence, which computers lack, to crowds of individuals who are willing to complete them cost-effectively. In different re-

---

<sup>1</sup>A Guide to the Web's Growing Set of Free Image Collections

search areas that require opinions, ratings, and labeled data, crowdsourcing has been leveraged to obtain large amounts of data with reliability similar to the data collected from other subject pools and panels.

As technology has become increasingly available worldwide, individuals have started to share similar ideas, contents, and aesthetic artifacts online. Publishers of these materials have started to compete for the attention of viewers. In terms of the social media context, photographers are competing for more followers of their work. This competition may motivate them to come up with more aesthetic photographs that may interest their audience and attract more followers. Owing to the technology available, such photographers can take many pictures during their photo shoots. An automatic composition evaluation system can help them in two ways. During a photo shoot, a real-time composition system can provide feedback and suggestions to the photographer [17]. Afterwards, the images can be grouped by composition or certain principles of composition such as *balance*. The images that satisfy composition guidelines may be chosen for upload.

The substantial amount of visual data now available has created challenges for scientists, too. Inspecting the data that is collected for hypothesis-testing can be a cumbersome task. For instance, in plant biology, the morphological properties of cells and the mechanisms behind them are a topic of interest. This requires an analysis of the spatial distribution of certain materials and the subcellular structures producing them. Visual cellular data presents opportunities to complete this task. However, both manual and semi-automatic analyses take a great amount of time, while inspection by the naked eye leads to inaccuracies in analysis. An automated system that segments the subcellular structures accurately and analyzes the spatial distribution of materials of interest can save significant time and effort while retaining accuracy. This thesis research studies the issue of managing visual data from three different perspectives.

## 1.1 Viewer Perspective: Compositional Aesthetics

Capturing beauty is a longstanding quest in the history of humanity. This quest has led to efforts to understand creativity and the production process of aesthetic artifacts. Such efforts can be observed in the daily lives of individuals from professional artists to amateurs. For instance, hikers might spend time taking pictures of scenery they like. A designer might attempt to come up with an aesthetically appealing and attention-grabbing design for a Web site or a graphic t-shirt. Therefore, there is a need for clear-cut rules regarding how to construct good compositions in a controlled manner.

To develop these rules and understand their mechanisms, psychology, psychophysics, and art experts have conducted studies with human subjects [2, 3]. Fine arts experts have compiled aesthetic guidelines drawn from the careful observation of aesthetic works and the recognition of patterns among them [18]. Computer science has joined this quest, seeing it as an interesting application of artificial intelligence. The question, “How can we make computers appreciate aesthetics?,” has garnered significant attention. Data-driven approaches have improved a computer’s ability to provide numerical values for images’ aesthetic quality [19].

One of the aforementioned guidelines is balance in composition. A balanced composition can be described as a composition whose visual elements together

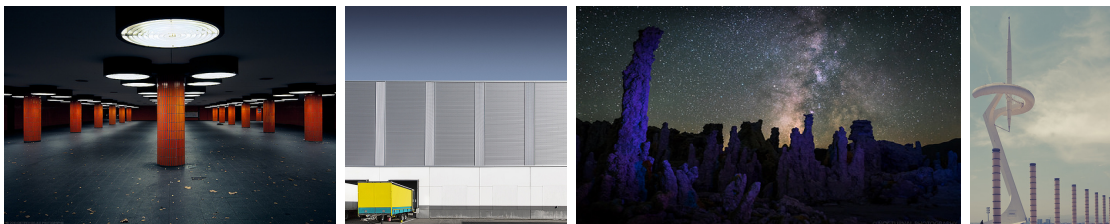


Figure 1.1: Balance in composition is an important aesthetic concept. It can be employed in different ways to have different visual impacts on viewers. The first image shows an example of symmetric balance, whereas the second image is imbalanced. The third image is imbalanced due to the tree trunk. The fourth image is balanced, as the fence is leading the eye away from the tower.

achieve visual equilibrium, whether the composition be a photograph or a painting. These elements include objects, textures, colors, and shapes, among others. In its simplest form, balance can be achieved by using pure symmetry along the vertical and horizontal axes of an image (Fig. 1.1, first image). An imbalanced image has the effect of creating tension and producing uneasy, disquieting responses in viewers. For example, the purposefully placed yellow truck in the second image in Fig. 1.1 draws our attention to the lower-left corner. However, there is another type of balance that is much more intriguing. The experts call it “dynamic balance” or “asymmetrical balance.” This type of balance occurs when the area of a certain visual element is balanced with another element with different visual properties [18]. The tree stump and Milky Way, or the tower and series of poles, demonstrate this phenomenon. Balance is a highly subjective concept, meaning that different observers of the same photograph may have very different views. This subjectivity makes modeling balance highly challenging.

The concept of balance plays an important role in computational aesthetic studies of the automated aesthetic layout of magazines and Web pages, automated image cropping and retargeting, seam carving, and aesthetic quality assessment [20, 21]. The representation of balance has been primarily calculated through the center of mass (CoM) of a saliency map [22]. A well-studied topic in computer vision, the saliency map of an image indicates those regions of the image to which humans pay more visual attention. Different methods have been proposed to predict the saliency map of an image.

It seems natural to employ saliency maps to represent balance in composition, as saliency maps and balance are both related to visual attention. In this dissertation, we challenge the assumed relationship and investigate whether there are other elements that can improve the representation of visual balance for computational aesthetic systems. This relationship is bridged through an analysis of visual centers and saliency maps. To the best of our knowledge, no study to date has created a larger dataset compared to those used in empirical arts studies, where the relationship among aesthetics, visual balance, and visual elements has been investigated through subject studies with small sets of a few hundred images. Therefore, we created an image dataset by compiling images from a popular photograph-sharing Web site. This dataset demonstrates different balance characteristics, as



articulated in art literature. The visual centers of the photos were obtained from individuals' responses to an online survey designed in line with empirical art studies [23]. Linear models were employed to measure the predictive power of saliency. The findings indicate that the predictive power of the linear model improves if the system accounts for the visual weight of humans in the image and the dominant vanishing point (perspective point).

## 1.2 Viewer Perspective: Emotions

With the advent of the information age, our lives have changed dramatically. The way people do things has shifted. This has led to a new learning culture known as “participatory culture” [24]. This culture is the result of lowered barriers for accessing expert knowledge. The new generation of learners is enthusiastic about collaborating on and solving problems. The Internet has eliminated the limitations of time and space. Thus, people around the world have the ability to participate in projects and achieve great things together. In return, organizations have altered their problem-solving approaches. Crowdsourcing has emerged as one such approach. There are different definitions of crowdsourcing that capture certain aspects of the phenomenon. [25] provided an overarching definition based on a comprehensive literature survey: *Crowdsourcing is a type of participative online activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task. The undertaking of the task, of variable complexity and modularity, and in which the crowd should participate bringing their work, money, knowledge and/or experience, always entails mutual benefit. The user will receive the satisfaction of a given type of need, be it economic, social recognition, self-esteem, or the development of individual skills, while the crowdsourcer will obtain and utilize to their advantage what the user has brought to the venture, whose form will depend on the type of activity undertaken.*

The rise of crowdsourcing has also led to changes in research methodologies, as crowdsourcing allows researchers to reach a fairly large participant base with diverse knowledge backgrounds in a short amount of time. Experiments, data-labeling tasks, and other small aspects of research (microtasks) can be crowd-

sourced. This type of crowdsourcing is considered the typology of the distributed human intelligence task (HIT). Apart from HITs, online collaborative communities where participants evaluate and contribute to one another’s work can be considered crowdsourcing, as the opinions and expertise of the masses are leveraged. For instance, recent computer vision datasets that are related to affect and aesthetics are being compiled from online collaborative photography sites or through HITs posted on associated platforms [26, 27]. Similar approaches have been adopted in affect-labeled audio, video, text data acquisition, and live map information-gathering [28, 29].

As more and more researchers come to utilize crowdsourcing for data, questions regarding the quality of the data and participant base have been raised. Studies have been conducted to assess the reliability of the data, analyzing participant demographics and comparing the data to data gathered via other, more traditional methods such as subject pools [30, 31]. Along these lines, although crowdsourcing has been utilized for visual data with affective and aesthetic preference ratings, the affective status and aesthetic preferences and biases of the crowdsourcing participant pool have yet to be inspected. In this dissertation, we attempt to shed light on the biases and differences in affect evoked by visual features frequently used in affective computing and computational aesthetics across latent demographics groups. Better surveys will be able to be designed using the insight into biases and demographic differences that is gleaned from our study.

## **1.3 Plant Biology Perspective**

### **1.3.1 Automatic Segmentation of Subcellular Structures**

In biology and medicine, tubular structures have important roles. Located at various levels of organization, tubular structures allow for the transportation of fluids or other biological structures. Examples at the system level are blood vessels in animals and xylem and phloem in plants. At the cellular level, microtubules (MTs) may act as rails that guide other subcellular structures [32, 33]. One such relation can be observed in plant cell growth.

The mechanisms behind plant growth and morphological development have

been a topic of interest for plant biology researchers. Cellular anisotropic enlargement (expansion) is achieved through a combination of outward water (turgor pressure) and non-uniform resistance from the cell wall.

As the membrane is covered by the cell wall, expansion requires certain mechanisms to be in place in order to prevent cell explosion. [34] indicated that growth is achieved via an increase in cell volume caused by water intake to the cellular compartments (*vacuoles*) and the extension of the pre-existing cell wall. However, the thinning of the cell wall owing to this extension must be compensated for by the integration of new polymers into the wall in order to constrain lateral cell expansion. Cellulose microfibrils transversely cover the cell wall in combination with a matrix of complex carbohydrates (*polysaccharide*), preventing the thinning of the wall, as seen in Fig

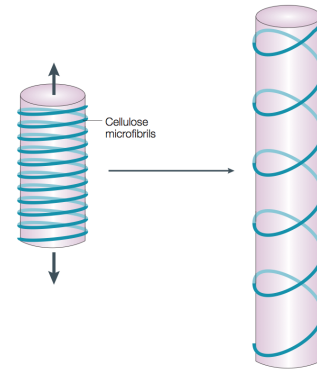


Figure 1.2: Microfibrils transversely cover the cell, constraining lateral growth [1].

1.2. Cellulose synthesis is accomplished by cellulose-synthesizing complexes that consist of cellulose synthase (CESA) proteins. CESA proteins are found in the plasma membrane as hexameric rosettes complexes [35].

The distribution and orientation of CESAs influence the deposition of cellulose in different parts of the cell wall. These parts of the cell wall may develop different morphologies, giving way to various functionalities. [36] showed that these particles are “motile” and traced in approximately linear trajectories. One model that attempted to explain these trajectories was proposed when microtubules were discovered. According to this model, microtubules set a template for cellulose synthesis and CESAs follow their orientation. Researchers have been investigating the colocalization of CESAs and microtubules to justify the model and further understand the inner workings of the cellulose production and deposition processes [33]. This presents new research opportunities for bio-image analysis.

An interesting instance of anisotropic morphological development is stomatal guard cells. Guard cells play a fundamental role in the gas transfer required for respiration, photosynthesis, and transpiration (the water movement process within

plants via evaporation) by bordering stomata (pores). The opening and closing of stoma is mainly caused by turgor pressure along with differently sized depositions of cellulose and matrix polysacchride in different locations. Further study of the interactions between CESAs and microtubules in guard cells, as well as the orientations of microtubules in open and closed states, might reveal more about the mechanics and functionality of stomata. The aforementioned research direction suggests the need to quantify colocalization analysis from microscopy images. Computer vision comes into play at this point. In [37], a rather detailed review of methods for colocalization analysis is provided. Image analysis tools based on different approaches such as correlation-based and object segmentation-based colocalization have been proposed by researchers [38, 39]. Each approach has its own merits and shortcomings. Nevertheless, object segmentation methods have long been a focus of attention in computer vision research, and some researchers would argue that this is an established area of study. A segmentation system can be used to distinguish subcellular structures that experts cannot. It can also process a huge volume of microscopy data in a shorter amount of time than experts can.

Computer vision research aims to help medical doctors and biologists by automatizing the detection of tubular structures. As there are different imaging modalities for different research purposes, there have been many different methods used in an attempt to solve the same problem. Some methods are based on models or fixed filters [40, 41, 42, 43, 44]. Some studies have approached the subject from the active contours perspective due to active contours' topological advances [45, 46, 47]. Recent developments in neural networks research have also been employed to solve the problem. Machine learning has been used in cases where researchers have sufficient data.

Different fields use 2D and 3D imaging data for different purposes. Computer tomography and positron emission tomography (PET) are common 3D modalities in medical imaging. Retinal scans and angiographs are acquired in 2D. In plant biology, demonstration of microtubule analysis is often done using 2D maximum intensity projection images [48]. These images can be challenging for manual analysis and computerized methods due to (i) the high number of microtubules,

(ii) contrast loss because of the tubules' position on the imaging plane, and (iii) tubules crossing over one another or the inter-tubule distance becoming shorter than the resolving power of the microscope.

### 1.3.2 3D Surface Extraction of the Stomatal Complex

3D models of biological structures can be utilized to simulate and analyze these structures' responses to different conditions and external impacts. Furthermore, reliable 3D models can provide us with more details about the physical dynamics of a specific complex. One area in plant science that needs such models is stomatal research. Due to stomata's aforementioned role in the gas intake of a plant, understanding underlying mechanisms in their opening and closing might yield crucial knowledge for different real-world applications such as irrigation planning or the design of industrial plants resilient to drastic changes in the rainfall regime.

The methodology used for stomatal analysis generally follows a common pattern across different studies. A hypothesis regarding the relationship between a function or property and a basic structure such as a gene or substance is proposed. The samples from the model plant family that show differences in the expression of the gene/substance under investigation are separated into mutant and control groups. The samples are imaged using an appropriate method, which is usually confocal fluorescence microscopy. The characteristics of the function or property under investigation are manually analyzed and quantified and the differences between the sample groups are statistically identified. Physical properties such as stomatal pore dimensions and area, whole stomatal complex dimensions, single guard cell dimensions, arc length, and shape play an important part in this analysis. An automatically or semi-automatically extracted 3D model of the stomatal complex might be used to accelerate this process. Such a computerized model might be more accurate in its measurements as it would be better able to represent the smallest topological changes than a manually crafted model could.

3D model extraction is related to the segmentation of relevant surfaces in a volume of interest. This is a well-researched area of the computer vision field. In particular, active surfaces are widely employed in medical image analysis in order to isolate the brain, other organs, and skeletal structures. Their use in cases of

brain segmentation is indicative of their ability to handle topological changes and gaps on surfaces. We devise a software program that utilizes the flexibility of both 2D active contours and 3D active surfaces to extract a 3D model of the stomatal complex. The challenges of completing this task are

- contrast loss deeper into the sample causes gaps in geometry,
- the surrounding pavement cells make it difficult to detect inherent stomata walls,
- the inner and outer surfaces need to be distinguished from one another for more accurate analysis.

In light of the discussions above, we propose a novel method that is adaptive to microtubule topology and that connects broken parts in their segmentation while adhering to their geometry. We come up with a framework that integrates a stretching open active curve system and a tensor voting framework. Our proposed method fills the small gaps where the continuity of tubules is broken due to different imaging factors. This work is followed by development of a software package to obtain the 3D surfaces of a stomatal complex through active surfaces and polar snakes.

## 1.4 Problem Statement

We propose four research branches addressing the issues mentioned above. The first one is an automated compositional evaluation tool with the capability to modify images. The second is an investigation into preferences and biases of crowdsourcing demographics, followed an automated visual analysis tool for microtubules at the subcellular level. Finally, the fourth is a semi-automated system for stoma guard cell pair isolation in 3D.

### 1.4.1 Research Questions

Our research questions focusing on the user-adaptation, compositional evaluation system, and the visual analysis of stomata cells can be categorized into two groups.

The first group is related to aesthetics and affect, and their relationship to visual features. We first consider how well saliency maps can predict visual balance and how this prediction can be improved. In relation to this research question, we seek to identify whether visual features commonly employed in computational aesthetics and affective analysis elicit any difference in response among various demographic groups in relation to emotional constructs and preferences.

The second group is composed of two main parts. The first inquiry that we make in this second group is into how active contour based methods can be further improved to follow tubular structure geometry. The second inquiry is about extraction of the 3D surface of the stomatal complex in the face of imaging imperfections via active surfaces.

### 1.4.2 Contributions

Our contributions in light of the aforementioned research questions can be summarized using the same groups. Our contributions for the first group of research questions can be summed up as follows. We created an image dataset that shows different balance characteristics, which are symmetry, asymmetrical (dynamic) balance, and imbalance, with visual center votes from participants, and we identified visual features other than saliency that may be related to visual balance. The relationship between saliency maps and visual centers, or visual balance, was inspected and it was shown that including high-level features such as human locations and vanishing point detection improves performance. A dataset including 41,255 images and the affective responses, aesthetic ratings, and demographic information of 2,063 participants was compiled. We showed that the participants could be separated into different latent demographic groups and that there were differences across these demographic groups in terms of affective and aesthetic responses to the images.

Contributions in relation to the second group of research questions can be summed up in three points. First, the initialization of snakes for the segmentation of microtubules in stoma guard cells was heuristically improved. Then, tensor voting was integrated into the stretching open active curve framework to obtain

better structural awareness. Finally, in a similar vein, an active contour-based method was devised to extract the 3D surface of the stomatal structure.

## 1.5 Structure of Dissertation

The dissertation proceeds as follows. In order to give reader some context for the work detailed in the dissertation, Chapter 2 and Chapter 3 are dedicated for background and related work. The former provides background research on art theory and the psychology of emotions and composition from a computational perspective. Compositional methods utilized to extract those visual elements that facilitate the creation of a balanced composition are laid out, and a discussion of how balance is used in multimedia is provided. In addition, a discussion of how crowdsourcing is utilized in computer science is given. The following chapter provides background on segmentation and denoising methods that are used for tubular structure detection and 3D surface extraction. The remaining chapters provide details about our research. Chapter 4 expands on the representation of compositional balance in images, and it is succeeded by the chapter that describes demographic differences among participants in our crowdsourcing pool. Following two chapters focus on biomedical image processing research. Chapter 6 discusses a novel method of subcellular microtubule segmentation. Ensuing chapter introduces our software package for 3D surface extraction. Last chapter discusses the implications of our work and the future research that is possible.



# Chapter 2

## **Background: Computational Aesthetics and Affective Computing**

### **2.1 Introduction**

In this chapter, some context and research background for our studies related to viewer perspective are delivered. The chapter cover research from various fields such as empirical art, psychology, and computer vision. By doing so, we aim to build a bridge between the knowledge base in arts and psychology, and computer vision. Relationship between aesthetics, emotions and visual features are provided.

We first introduce compositional balance as an element of aesthetic. A discussion on what type of visual cues are used to strike compositional balance is given. The following sections introduce emotions and relationship between emotion and preference. We continue on with phenomenon of crowdsourcing, and how it is utilized in computer science. We lay down why we are interested in demographics groups and provide information on personalization.

### **2.2 Aesthetics**

Empirical studies of aesthetics and perceptual psychology have sought to lift the veil on aesthetic phenomena. Under laboratory conditions, researchers modify visual elements (different shapes with different colors in diverse compositions) and

seek participants' aesthetic opinions of them. [18] documented some of the common practices of art. The Gestalt school of thought in art looks for these common practices in an overall image and considers the design in its entirety [23]. These common practices include (i) *unity*, which is the practice of putting elements of a picture or a design together in order to give a sense of its entirety; (ii) *emphasis*, which is using elements to draw visual attention to some part of the design (picture); (iii) *scale*, which is the implementation of objects or patterns at different sizes so that the 2D picture gives a sense of three dimensions; (iv) *rhythm*, which is the use of repeating patterns to produce a desired effect in viewers; and last but not least, (v) *balance*, which refers to the equilibrium of elements in the artifact.

Of course, artifacts of interest can also be seen as data to be processed. In this case, regularities in the data may be justified and new guidelines are necessary. Thus, computer scientists have become interested in the problem of aesthetics as computing power has expanded. Computing power could help artists, amateurs, and designers to create better artifacts with ease and speed. This line of research has the potential to refine information retrieval systems. An image search might return more aesthetically pleasing results through computer vision's application of aesthetics theories to information retrieval [49]. There are applications that can evaluate the aesthetics of Web page designs and automatically generate magazine covers [50]. In considering an avid social media user with a low level of photography knowledge, it seems that a system that provides suggestions for a better composition that are tailored to the user's demographics could be useful because such a system would help the user to express their own point of view of an object or a scene. Research on automatic composition has garnered significant attention, as it has many applications in fields such as publishing and photography. Before making any effort to identify algorithms, however, it is important to have a solid understanding of relevant theories and discussions in the fields of art, empirical art society, psychology, and physiopsychology. One important step in this process is grasping the elements and observed principles in design. Of the aforementioned design practices, balance is the most relevant to composition. A deeper understanding of balance could lead to design of better algorithms to quantify balance.

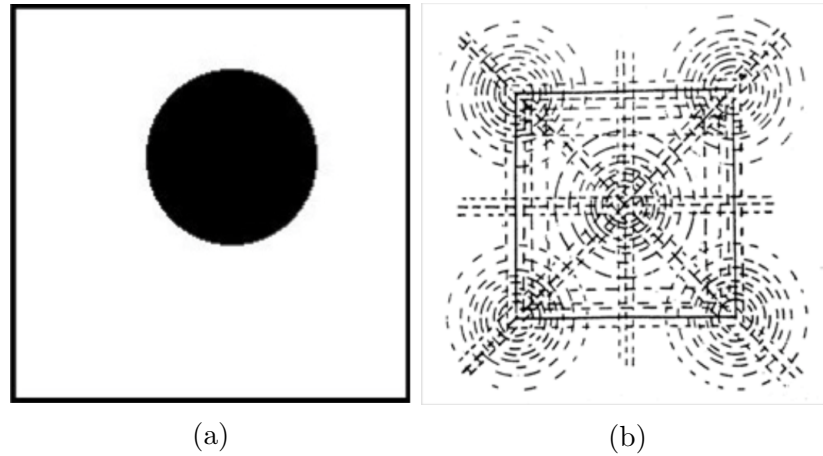


Figure 2.1: The off-center disk in a square frame and the so-called structural net with equilibrium points [2].

## Compositional Balance

As previously mentioned, aesthetics and its elements have been a topic of interest among art scholars. A full discovery of elements could lead to a formulation for aesthetics. Gestalt principles define balance as *the equilibrium of visual weights in the design or artifact*. Scholars who adhere to the Gestalt school [2, 51] consider balance a prominent part of aesthetics. In [2], an explanation of balance is given using a black disk with a square frame around it, as seen in Fig. 2.1a. The existence of a perceptual field of push and pull like a force field within in the square frame is mentioned; see Fig. 2.1b. The eye assigns an object to an equilibrium position within the push and pull field. If the object is not located at one of these equilibrium points, the composition feels disjointed to the eye. Thus, the author claims that balance is one of the factors of aesthetics. The author does not provide any testing scheme for these theories, however. This has led to conflicting results in subsequent studies that have sought to prove the hypothesis.

The first conflict is related to the definition of balance. The literature reveals concerns about the ambiguity of Arnheim's definition, as researchers are not sure whether the equilibrium described is metaphorical or physical. The approach that most researchers have taken is to view the equilibrium as physical. Thus, researchers have attempted to prove the aesthetic theories of balance from the physical perspective. Even adopting this same perspective has not led to a

unity of results, however. Rather, studies have returned conflicting results. In [3, 23, 52, 53], results opposing the relationship between balance and aesthetics were reported, whereas Arnheim’s theories were supported in [54, 55, 56].

In [23], participants were asked to put a fulcrum under “paintings of accepted merit,” using both the original and cropped versions. It turned out that most of the paintings were not balanced in the center of their frames. The researchers also reported that the relationship between objects’ size and distance was non-linear to viewers, unlike the physical relationship, which was linear. The issue was rekindled in [3] by conducting a similar experiment. This study looked at balance from the same physicalist perspective used by Arnheim. A few experiments were conducted in an attempt to determine the accuracy of this proposed model. Artistic photographs and random photographs were compared. First, the center of mass (CoM) of each photograph was calculated. According to these calculations, the CoMs of artistic photographs were aligned with the axes of the images more frequently than they were in the random photographs. The CoMs of artistic images were mostly aligned with the central vertical axis. This finding seemed to affirm Arnheim’s equilibrium principle. However, the other experiment yielded conflicting results. In the second experiment, the authors shifted the CoMs by modifying the gamma of the pictures, which is related to the contrast and intensity of images. Each CoM was moved off (to the left or right) the central axis if it was aligned with the central vertical image or vice versa, as shown in Fig. 2.2. The participants were asked to aesthetically choose between the originals and the modified versions. The results were inconclusive as there was no obvious preference for images with CoMs aligned with the vertical axis. The third and fourth experiments focused on balance through the cropping of images. While the third experiment seemed to support Arnheim’s principles, the fourth experiment, which was more complex, showed no predominant preference for CoM-on-axis images. The fifth experiment was conducted using Arnheim’s black disks, and the results supported those of Arnheim. The conflicting results of experiments published as part of the same study demonstrates the level of ambiguity. In [52], the meanings of “aesthetic,” “balanced,” and “weight equilibrium” were investigated. The correlations among these terms were low. The authors pointed out that aesthetic ratings were high for symmetrical but imbalanced compositions and questioned how far an object

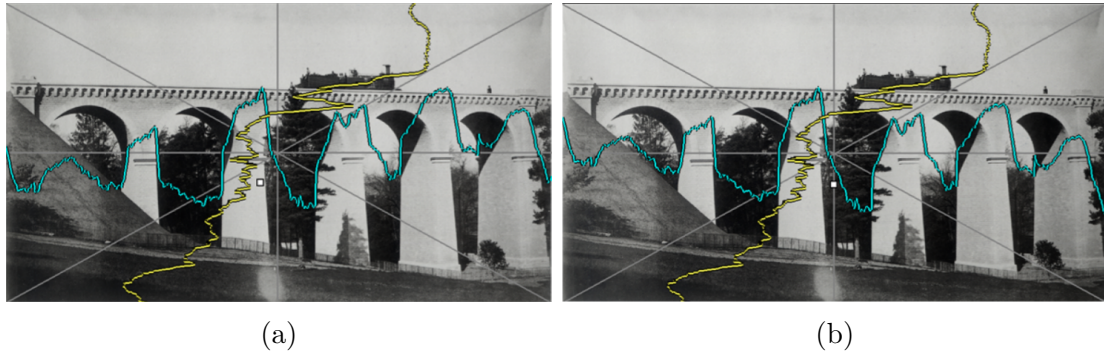


Figure 2.2: One of the original images in the study (a) and a modified version of it wherein the gamma ramp was applied (b). The white point represents the center of mass (CoM) [3].

should be placed from a different fixed object within a frame in order to obtain a good composition [53]. They found that balance had almost no influence on the aesthetic placement of the second object. Rather, the participants placed the second object centrally.

There are also opposing results in the literature. Participants were asked to design aesthetic compositions with black disks, squares, rectangles, and leaves of different sizes in [57]. The participants' design processes were photographed at various intervals. Most of the designs were balanced around the central axis and showed some sense of symmetry. The same sorts of experiments were conducted with similar results in [54]. As last two paragraphs indicate, the results of these studies vary. The problem with conflicting results demonstrates that researchers do not ensure participants can truly perceive the weight equilibrium. In [52], it was shown that participants miss balance even if the CoM is aligned with the center. Most of the experiments were conducted with abstract shapes instead of real-world pictures. Hence, the principles set out by Arnheim are still under exploration. Strangely, contradictory studies have failed to come up with a consistent replacement model. This supports the idea that computer scientists should continue to use Arnheim's principles.

Within this state of contradictions, we chose to follow Arnheim's principles because they are largely accepted in design and art books [18, 58]. Thus, we proceeded with the settled factors of balance within the Gestalt school's and Arnheim's frameworks. The design practices that are typically implemented to achieve

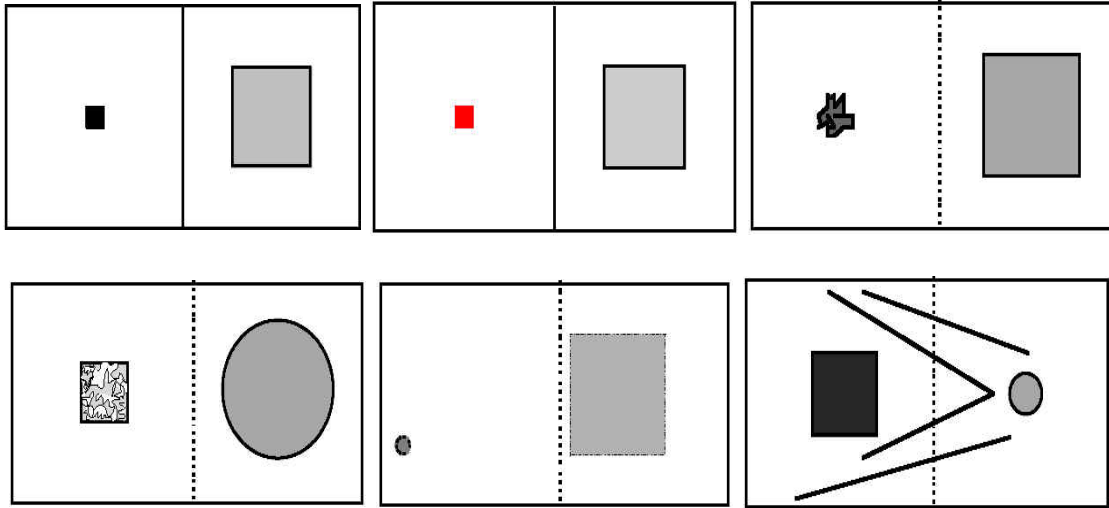


Figure 2.3: The elements that can be used in balance from upper-left to lower-right: value, color, shape, texture, position of object, and eye direction [4].

balance are symmetry, shape, color, texture, and the use of differently sized objects [18]. Axial symmetry's contribution to balance is obvious. The contributions and use of other elements are more subtle. The balance achieved through the use of these elements is called “dynamic balance.” Different structures with different densities are used around the central vertical axis. In [4], the visual elements used to strike a balanced composition were identified as value, color, texture, shape, position, eye direction, and physical weight, as shown in Fig. 2.3. Ideas related to visual attention follow, as high-contrast parts of an image draw more visual attention so they can balance the larger, low-contrast regions. Brighter colors of small regions can balance bigger regions of more neutral colors. A large plain area can have a similar weight as a small region highly cluttered with complex shapes. Similarly, a small, high-contrast texture area can have a weight equal to that of a bigger, smoother area. In terms of objects, objects with similar texture, contrast, and shape but different sizes can be positioned in such a way that the smaller object is closer to the boundary of the image, whereas the larger object is close to the center. In addition, directional lines (i.e., perspective elements) can be used to balance an image by drawing visual attention from one side to another.

At this point, computer vision comes into play, as symmetry, color segmentation, and texture segmentation have been well studied in the field. These methods are discussed further in the following section.

However, the appreciation of aesthetics is a multifaceted and complex event. There are cognitive and affective (relating to moods, feelings, and attitudes<sup>1</sup>, which are detailed later) sides to it. As cognition and affect are phenomena that are rather personal, comprehending the cognitive and affective workings of individuals is necessary to move closer to building personalized aesthetics systems.

## 2.3 Balance and Aesthetics in Multimedia

Interest in aesthetics has recently risen among researchers in the computational domain. The subjectivity of aesthetics is recognized; however, data-driven approaches have shown some consensus can be reached [59]. In this regard, there have been efforts to automatically make visual data more aesthetic or to provide feedback to users. These data-driven approaches have also been applied in other areas. In [26], an attempt was made to leverage aesthetics scores to learn a computational model to aesthetically rank images. The findings about aesthetic scoring were applied to measure the aesthetic value of portraits in [60].

In this research field, balance is considered a small feature to be included in some ad hoc methods and its definition is not very clear. In [22], the salient regions of an image were used to quantify the sense of balance in the composition. However, this analysis did not consider other factors such as objects or shapes. Other approaches that have been used are related to Web page analysis and rely on the spatial distribution of colorfulness (entropy) and Web pages' image structure using the XY-cut method [61, 62]. These approaches are based on the CoM and rely too heavily on low-level features. Balance is related to higher-level features, and the focus of our attention should be on this aspect. Bringing in connected components in a picture can be a helpful tool. A review of the literature focusing on how to extract features to capture balance is given in the following subsections.

---

<sup>1</sup>Oxford Advanced Learner's Dictionary

### 2.3.1 Symmetry

One idea of symmetry comes from [63], who explored how visual features can be employed in emotion recognition. This study made use of ideas put forth in [64] and [65]. In [64], the interest points in a scene were spatially transformed to detect radial symmetry. In [65], the feature points were grouped together and their local and global symmetries were checked. This allowed the researchers to use the spatial distribution of symmetry points to capture static balance. These algorithms' implementations are available online. Balance can be better quantified if we can:

- obtain the symmetry points, and
- look at how those points are spread around the image. This can be done by dividing the image into nine grids and calculating the number of symmetry points within these grids. This is considered a global sign of balance.

### 2.3.2 Texture

Texture plays an important role in composition, according to [66]. The features mentioned in the study were considered relevant to aesthetic concepts. In [67], these features were also used and added to the wavelet analysis. Hence, we propose to first calculate texture properties using:

**Gray Level Co-occurrence Matrix:** A matrix with entry at index  $i, j$  describes the frequency of pixel pairs with intensity  $i$  and  $j$ , separated by a displacement vector  $d$ . Contrast, entropy, and correlation are calculated based on this 2D histogram. These values are computed for different angles and lengths of  $d$ , where  $\{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$  and  $(|d| = 1, 2, 4, 8)$ .

**Tamura Features:** In [67], texture features of coarseness, contrast, and directionality were used in affective image classification. These features were selected from the list of features given in [68], who based their list on psychological experiments. The original list consists of coarseness, contrast, directionality, regularity, and roughness. The computational definitions of these features can be found in [68].



**Gabor Filter Output:** The classical sinusoidal wave with a Gaussian envelope is convolved with an image at different frequencies and orientations. The filter is good at detecting structures with different orientation; however, determining the frequency at which it works well is a problem. In order to detect texture differences, we can consider a range of frequencies and create a bank of Gabor filters. As the filters convolve with the images, each pixel can be assigned the orientation and frequency at which it gives the maximum response. Similar textures should have similar responses for similar filters.

**Fourier Spectrum:** The response can be used to determine whether the image is dominated by radially distributed texture. If it is, the frequency components will be almost equally distributed radially.

**Wavelet Analysis:** Wavelet analysis can be used to measure the smoothness and graininess of an image's regions. We can follow the same approach; however, this method is more effective if it is used with segmentation so that the textures of the objects or regions can be analyzed. The area of each texture class, or the area of each object, can be compared to the areas of the other texture classes.

### 2.3.3 Shape

For this property, the contour algorithm laid out in [69] can be used. In terms of contour extraction, their method basically fits a disk on a pixel and checks oriented gradients and color information to assign a contour confidence value. After this method is applied, two types of analysis can be used. The first one involves computing the CoM of the contour map. The second one involves obtaining a connected component analysis and looking at the distribution of the connected components in a nine-cell grid. The counts of connected components in each cell can be turned into a vector where the origin is the center of the image. If the shape to be detected is known in advance, the Hough algorithm might serve the detection purposes. In this method, the shape parameter equation is transformed into the related Hough space. Each point in this space represents the same shape with different parameters. Each point in the Hough space is cast a vote according to the number of edge pixels that each possible Hough shape covers in the image.

### 2.3.4 Color and Value

Spatial color distribution makes up an important part of the balance analysis. The first step is the naming of the colors in an image. There are two approaches to this, as seen in [70] and [71]. If we are able to obtain the dominant color names for the segmentation result, we can compare the area of one color to the areas of other colors based on previous research [72]. Balance between two colors is achieved if the colors are arranged in such a way that the area is in inverse proportion to the hue\*saturation. When working with natural images, this process becomes more complicated. The same concept was mentioned in [73] and the bag-of-color-words was utilized, but it was used for aesthetics and was not considered from a compositional point of view. The words were constructed by using local regions of simple color composition.

### 2.3.5 Depth

Using Make3D [74], we can obtain depth information and reconsider the segmented objects in an image. We can compute a CoM where the depth value is the mass. The closer an object is, the heavier it will be in visual terms. This method can be used to represent cases in which dynamic balance is achieved by objects of different sizes that are positioned at different depths.

While analyzing an image for dynamic balance, we need to take locality into account. A segmentation of the scene can be produced based on [75] or [76]. The features above can then be calculated for objects within the image. In order to convert object information into the visual tension field, we can represent the objects with their CoMs and related vectors.

### 2.3.6 Saliency

Salient regions in an image are the parts of an image that the eye spends more time on (“fixates on,” as is said in psychology) while viewing the image. According to a 2013 study, “Saliency intuitively characterizes some parts of a scene—which could be objects or regions—that appear to an observer to stand out relative to their neighboring parts” [77]. Saliency has been considered from two cognitive angles, which

are bottom-up and top-down. Bottom-up models focus on low-level properties, whereas top-down approaches focus on tasks and related heuristics. Approaches to capturing saliency can be categorized as follows (Table 2.1): (i) information theoretic models, (ii) cognitive models, (iii) graphical models, (iv) spectral analysis models, (v) pattern classification models, (vi) Bayesian models, and (vii) decision theoretic models.

Graphical models employ a graph structure to capture the independence among the random variables that represent the image. As there are latent variables reflecting saliency information, hidden Markov models and conditional random fields (CRF) can be applied. In [78], CRF was employed to learn what makes up salient regions from the data. Model complexities and interpretability difficulties are inherent to graphical models. Another approach that uses supervised learning is given in [79] by training support vector machines (SVM). However, a problem with testing learned models arises as the models heavily depend on the data.

The definitions of saliency that are accepted by researchers show differences in their purposes. Some studies have focused on segmentation [80, 81], while others have focused on visual attention [82, 83]. Naturally, a common theme is the measure of *dissimilarity*, but the studies take different approaches to it. In [82], the scientists did not present new features but instead speculated on a given feature map. The feature map was transferred into a fully connected directed graph and its edges were weighted proportionally to their dissimilarity to the surrounding region and distance from other nodes of the graph in order to compute an activation map. Based on this graph, a Markov chain was defined by “drawing equivalence nodes and state, and edge weights and transition probabilities.” The equilibrium state of the chain placed more importance on nodes with higher dissimilarity, yielding an activation map. The activation map was normalized again and transferred to a graph using a similar Markov chain scheme. This algorithm had a center bias, which may have adversely affected composition analysis.

In [81, 83], region information was included and a multiscale approach was utilized either by patches or by initial segmentation. In [83], context information was incorporated. Local and global information was combined to achieve this. A dissimilarity measure was created for a given patch by utilizing the Euclidean distances between color vectors of the patch and distances regarding patch positions.

Category	Name	Description
<i>Patch-based</i>	DRFI [81]	Dissimilarity measure defined among segmented regions in multi-scales.
	Context Aware [83]	Dissimilarity defined among patches with a spatial distance constraint for context.
	Achanta [80]	Frequency information through different channels leveraged for dissimilarity.
	CovSal [84] SWD [85]	Covariance among patches employed for dissimilarity information. Dissimilarity patches represented in reduced dimensional space utilized.
<i>Graph-based</i>	GBVS [82]	Pixel information such as color, intensity, and orientation incorporated into a graph, where the edges represent a dissimilarity measure. The graph is treated as a Markov field to obtain an activation map.
<i>Center-surround</i>	CorSal [86]	Corner cues obtained via application of Gabor filters used for saliency information.
	FES [87]	Bayesian framework between visual features and saliency within a moving center-surround window.
	Murray Model [88]	Center-surround filter size and other parameters learned through GMM.
<i>Others</i>	ImSig [89]	Saliency map based on inverse DCT of image signature for foreground-background separation.
	UHF [90]	Unsupervised hierarchy of visual features for saliency.
	MDF [91]	Multiscale feature learning through CNN for saliency inference.
	Shallow Convnet [92]	A shallow neural network followed by a deep network in order to regress saliency values.

Table 2.1: State-of-the-art bottom-up saliency methods summarized according to their approaches.

The same analysis was done at different scales and the results were averaged to come up with a number. This saliency was redefined by using a distance measure to attention foci to incorporate the context information. Context information can be useful for composition analysis. A similar idea was applied in [81] using different levels of segmentation from fine to coarse. The researchers also came up with a new feature to better distinguish the background. They extracted features within segmentation regions and compared them for dissimilarity. Instead of directly calculating saliency, they computed the features and employed a learning scheme through a random forest regressor with saliency labels from previously obtained data. The researchers termed this dissimilarity “regional contrast.” In [80], the purpose of saliency focused on segmentation. The very low and very high frequency components from a given image were eliminated to obtain more uniform regions.

Basically, the saliency of a pixel was defined as the norm of the difference between the mean intensity of the image and the intensity value of the pixel in the bandpass-filtered image. The method seemed to work well for images with a dominant object.

## 2.4 Emotions and Affect

Emotions are a part of a phenomenon called *affect*. Affect is defined as the psychological state (sense-feeling) of a person during sensational interactions with the environment [93]. Its representation, classification, and mechanisms are the focus of psychology. The study of affective phenomena cannot be considered in isolation from other phenomena within and around a person. Therefore, affect involves cognitive processes and is related to our research topic, which draws from the field of empirical aesthetics. Affect presents an opportunity to understand what makes individuals appreciate the aesthetics of an object. Understanding this will help artists to produce improved aesthetic artifacts.

Affect can be divided into three subgroups: core affect, moods, and emotions. Emotions are the most interesting of this group, as they are directed towards an object. A detailed discussion of emotions is necessary for the sake of our argument. Emotions' representation methods are especially important for this research as emotions affect the machine learning scheme and data-gathering. There is substantial literature on the representation of emotions. The first approach is derived from evolutionary knowledge. In [94, 95], the researchers argued that people developed discrete emotions with related facial expressions. In line with this argument, the categorization of emotions according to facial expressions was adopted. These facial expressions can also be observed in primates. Humans have eight discrete emotions including happy, angry, and sad. Recent and improved approaches to emotions are the dimensional and circumplex approaches [5, 96, 97]. In [5], a 2D circumplex model was proposed, as shown in Fig. 2.4. Emotions can be represented with the two bipolar dimensions of “affective valence” and “perceived activation.” Another approach proposed is the Positive Affect(Activation)-Negative Affect(Activation) Schedule (PANAS) [96]. The dimensions in this approach are not bipolar, according to the authors. Rather, one dimension (PA) covers the

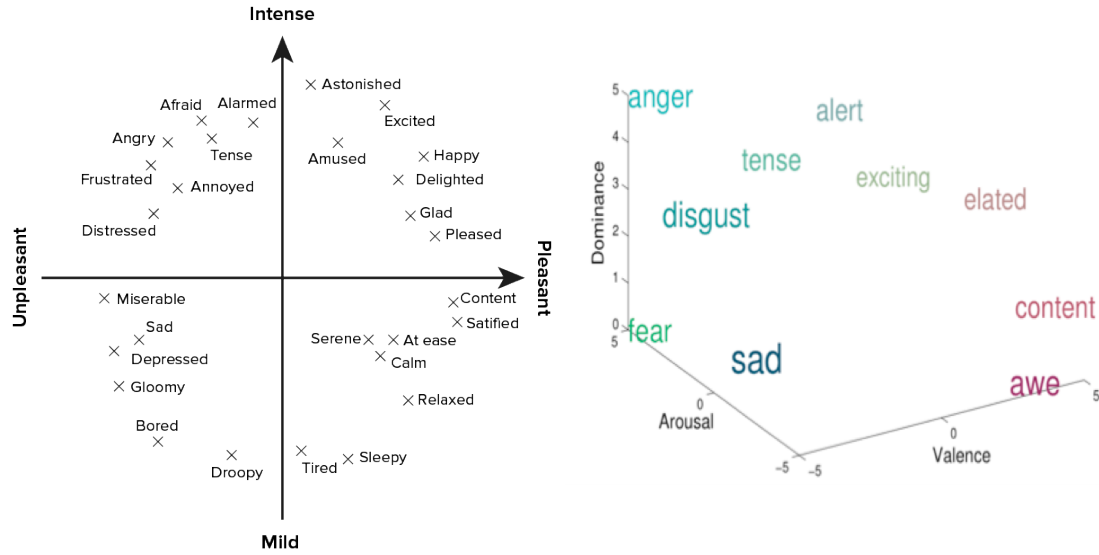


Figure 2.4: 2D circumplex proposed in [5] and 3D emotional space [6].

affective space from “high-activation pleasant affect to low activation unpleasant affect.” The other axis (NA) is from “high-activation unpleasant affect to low activation pleasant affect.” Confusingly, PA and NA sound as if they are bipolar in the definition; however, they are affectively unipolar. This means that only the state of high activation has affective value. Therefore, the dimensions are only represented by the high-activation poles. Another dimensional approach was proposed in [97]. The poles are defined according to arousal types. The first dimension is described as *the energetic arousal* (EA), which spans energy-tiredness. The other dimension is described as *the tense arousal* (TA), which spans tension-calmness. The authors stated that PA and NA are compatible with EA and TA, respectively.

The cognitive process of appraisal plays an important role in the eliciting of emotions. Appraisal was defined in [98] as “cognitive evaluations of events that are considered to be the proximal psychological determinants of emotional experience.” Certain permutations of appraisals match certain emotions. An appraisal takes place in two parts. Individuals consider an event’s or an object’s motivational relevance and motivational congruence. Motivational relevance is related to the intensity (arousal) of emotion. The more important an individual considers something to be, the more intense emotions the individual has. On the other hand, motivational congruence is linked to valence. Evaluated coping potential is

also associated with feelings of pleasantness or unpleasantness. If a person finds an object or event less manageable (copable), they have negative feelings. This has implications for arts and aesthetics. If an artifact is appraised as manageable, it appears aesthetic to our eyes. Balance may influence manageability. [99] described a study in which subjects were shown some samples from the International Affective Picture System (IAPS), which contains emotionally labeled pictures [11]. One group of individuals labeled images emotionally according to the valence and arousal dimensions; the other group provided information about their appraisal of the images. The two groups labeled those images according to their emotions and appraisals separately. It turned out that the emotional labels given by one group could be predicted by the appraisals provided by the other group. According to the authors, high-valence pictures received high ratings on the appraisal scale *agreeable/well-being*. On the other hand, *chance/human agency* (natural-ity) influenced arousal along with the *novel/unexpected* composite. At the end of the study, the authors stated that these appraisal processes could occur sub-consciously. They named them “cold appraisals.” What these results point to is important. By discovering more about appraisal scales and the visual properties that lead to them, new art and design principles can be explored. This might pave the way for a better understanding of art and, in the end, a better form of art. Empirical aesthetics research has also focused on the relationship between aesthetics and emotion. There are efforts from both the emotions front and the aesthetics front. Scientists have sought to come up with an accurate model to show the mechanisms behind aesthetic appreciation and judgment. In [100], the researchers explored the relationship between the feeling of interest and appraisal. Interest was presented as an emotion that resulted from an appraisal. The study showed that appraised complexity and appraised ability to understand a picture influenced interest. Individuals were more interested in a picture if the picture was considered more complex and understandable. In [101], an attempt was made to present a model of aesthetic appreciation and judgment, and the interest was in cognitive processing of art and how it elicited positive feelings. When an individual was shown an artwork, it involved “classifying, understanding, and cognitively mastering the artifact.” The model put the perceptual analysis first in the processing line. This includes basic occipital visual processing [101]. During this

stage, contrast was an important factor. The sharpness of an image increased the efficiency of the information-processing stage. Visual complexity was another important factor in visual processing. Adherence to Gestalt principles impacted information-processing mechanisms.

## 2.5 Emotional and Preferential Differences Across Different Demographics in Crowdsourcing

Different research fields have utilized the power of crowdsourcing. We focus on how crowdsourcing is employed for computer science, and what benefits we can gain by inspecting the *crowd*.

### 2.5.1 The Utilization of Crowdsourcing for Labeled Data

Various branches of computer science have employed crowdsourcing to obtain a considerable amount of labeled data. In the text retrieval field, for example, crowdsourcing has been leveraged for relevance evaluation, relevance feedback, and topical judgments [102, 103, 104, 105]. Crowdsourcing has also been employed for different language-related tasks such as the transcription of spoken language, reading, and interpretation [28, 106].

Computational aesthetics and affective computing studies have recognized the benefits of crowdsourcing and have leveraged it for the acquisition of labeled data. Affective labeling of multimodal media through Amazon Mechanical Turk (AMT) has been a popular method [29, 107, 108, 109]. AMT and other similar online platforms have been used to gather information about individuals' aesthetic preferences. There are datasets that include images crawled from community sites such as Flickr<sup>®</sup>, DPChallenge.com, and photo.net. One of the pioneers in collecting data using this method is [19]. The authors of that work utilized photo.net in order to collect photos with aesthetic ratings. Following this, they used a machine learning scheme to train the system with labeled images. In [26], DPChallenge.com images were used to create a dataset for visual aesthetic analysis. [110, 111] utilized Flickr<sup>®</sup> images and tags. In [112], Flickr images were collected and rated on AMT.



## 2.5.2 Amazon Mechanical Turk (AMT)

One of the most prominent crowdsourcing platforms is AMT. It was designed to have individuals accomplish microtasks that are complicated for computers. It is considered an inexpensive and rapid way to collect data [113, 114].

The AMT community is divided into requesters and workers (Turkers). Although some are skeptical about the data quality of AMT [30, 115], multiple studies have shown that the data quality is similar to or better than online panel participant or student subject pool data [113, 116, 117]. In addition to analyzing the reliability of the data, these studies provided insight into different aspects of participants such as their demographic distributions, motivation to participate, and the settings in which they take the surveys. It is estimated that there are more than 100,000 AMT workers. However, the literature suggests there is a discrepancy in the ratio of participants from the United States and from outside of the United States (mainly India) [30, 31, 113, 116]. According to [116], 47% of the participants are from the United States, and one-third are from India. Women make up 57% of the participants. The participants are from 50 different countries, including all 50 U.S. states. Non-American workers make up 31% of the population, and 36% are non-white [113]. A recent study showed that the ratio of non-U.S. Turkers to U.S. Turkers was much higher at 7:1 [30]. International Turkers are more educated than U.S. Turkers, although they also have lower incomes. [31] conducted a longitudinal study to better understand the demographic make-up of Turkers rather than relying on a snapshot in time. They confirmed that there are more than 100,000 subjects in the AMT pool; however, the population is more dynamic than expected where the workers have a survival rate (half time). A worker becomes obsolete after a while. They reported that Turkers from the United States make up of 75% the population, while Indian Turkers make up 16%. While they pointed to a younger worker population for Indian Turkers, they noted that Indian Turkers also have lower incomes on average than the general U.S. population.

In [116], workers were compared to community participants and college students by giving them the same questionnaire. It turns out that the AMT worker sample matched community samples from an age and gender perspective, but the AMT workers were older than college students on average. When it comes to their

motivation for participation, multiple studies have demonstrated that the workers are motivated by financial compensation rather than interest in the topic or a feeling of achievement [25, 30, 113, 118].

### 2.5.3 Studies on Differences Across Various Demographics

**Emotional Differences.** Emotional gender differences can be investigated through body measurements. This type of research looks into the activation regions in brain in fMRI [119], skin conductance measures [120], and facial electro myographic (EMG) activity when participants are presented with the same visual stimuli. In most of the cases, the International Affective Picture System (IAPS) [11] has been employed. These studies have reported the neurological responses of men and women to positive and negative affect stimuli differ.

**Preferential Differences.** The focus of preferential studies is naturally color. For different purposes such as psychology and website design [121, 122] and marketing and product design [123], researchers have analyzed responses from both genders to relevant artifacts. Psychological studies have attempted to find color preference differences between genders. The main approach employed by these studies is the exposition of a discrete color wheel or chart, and using a questionnaire about the wheel [124].

### 2.5.4 Personalization

As stated in Chapter 1, people who use the Web to share their ideas compete with other content-sharers; hence, they attempt to be different from others to reach their audience. This includes adjusting their content to fit the profile of users so that the users like what content-sharers are publishing online more. This topic has recently garnered attention in Internet sciences.

The personalization of Web pages has been increasingly investigated, as its importance is grasped by the Web community. We might need to consider the external elements that influence a user, since personalization is the process of tailoring a system to a user. These external elements can be the ones that are directly related to the person, usage information, contextual elements, and social elements [125]. In our study, we are interested in elements that are directly related to users,

which are demographics. The demographic context shapes how individuals appraise aesthetic elements. Hence, demographics are related to perceived usability through the discussion of appraisal given above. In [126], the adaptivity of Web pages to different cultures was inspected in detail. The researchers used cultural information to create a user model specific to a culture and adapted the interface according to this user model based on a culture ontology. This culture ontology was obtained through a combination of different elements. Some of them were gender, age, religion, education level, computer literacy, and language. Interface aspects were mapped to cultural dimension scores and user interface adaptations were suggested for these aspects according to users' cultural dimension scores. Studies that tied cultural influence to Web page design were conducted by [127, 128]. They utilized the local and foreign Web pages of a famous consumer electronics brand. The trust and preferences for both Web pages were measured for participants from different countries such as the United States, Canada, Germany, and Japan. It was shown that design preferences for local site were the most different between the Western countries (United States, Canada, Germany) and the Eastern country (Japan). Within the same cultural group, only Americans were loyal to the local site. This study raises the possibility of demonstrating cultural differences; however, it was conducted using only one Web page. Moreover, biasing might have been a problem due to the fame of the company.

## 2.6 Summary

In this section, the concepts that are essential to the dissertation research were introduced. As it is important to understand concepts before starting to do anything, familiarizing ourselves with the concepts of aesthetics, composition, and emotions was crucial. These concepts were introduced and certain relations among them were laid out. The principle of compositional balance was introduced, and the conflicting views about it in the empirical arts field were described. Certain visual features and their computational methods were also discussed briefly. Crowdsourcing, a current trend in ground-truth collection, was introduced alongside its different applications in computer science. The topic of personalization was briefly described.

# Background: Tubular Structure Analysis in Bioimaging

## 3.1 Introduction

In this chapter, image processing and computer vision techniques relevant to our study are discussed. As stated in Chapter 1, the tubular subcellular structure analysis of guard cells may benefit from medical image analysis research. Hence, we provide a review of the literature on image denoising and vessel segmentation in detail along with a discussion of useful morphological operations. The discussions about active surfaces are also provided in this chapter.

The chapter starts with brief coverage of stomatal research in plant biology. The following section briefly introduces morphological operations that we frequently employed for pre-processing and post-processing in our segmentation algorithms. We continue on with relevant denoising research. The ensuing section provides information about different algorithms for tubular structure detection. Finally, we give information on the tensor voting framework.

## 3.2 Stomata Research in Plant Biology

Before going into technical detail on computer vision aspect of stoma research, a short background is provided here to give context to the problem at hand. We are

interested in two particular aspects of stomatal research such as microtubules as subcellular structures and 3D surface modeling.

### 3.2.1 Microtubule Research in Plant Biology

Stomata play a fundamental role in gas exchange between the atmosphere and a plant. A pair of guard cells borders a stomatal pore, and the opening and closure of the stomatal pore is mainly caused by changes in turgor pressure in the guard cells. Dynamics, production, and distribution of microtubules (MTs) and their relations to stomatal mechanics have garnered interest among researchers in the plant biology community. Most of these studies depend on counting, tracing, and quantification of MTs for different states of stomata, all of which are usually manual processes [129, 130, 131]. Colocalization analysis of MTs and CESAs benefits from MT tracing and quantification. [37] provides a detailed review of methods for colocalization analysis. Most methods of colocalization analysis that have been proposed are either correlation-based or object segmentation-based [38, 39]. Nevertheless, object segmentation methods have long been a topic of discussion in computer vision research, and one may argue that this approach has an established foundation. A segmentation system can be utilized to distinguish among subcellular structures that experts cannot. Such a system can process a huge volume of microscopic data in a shorter amount of time than that needed by an expert.

### 3.2.2 3D Surface Modeling

Plant biologists are interested in 3D models of stomata as they can use them simulating different scenarios or explaining certain behavior. The models can also be utilized to compare against research results from different studies. In [12, 132], physical aspect of stomata are investigated, and 3D modeling is employed for analysis. The physical modeling requires parameters such as elasticity, pressure, tension in order to predict behavior. The relationship between parameters and actual behavior can be learned by matching visual stoma data to those parameters. If the visual data is more precise, modeling becomes more accurate. Visual data can be better represented via an automated surface extraction method.

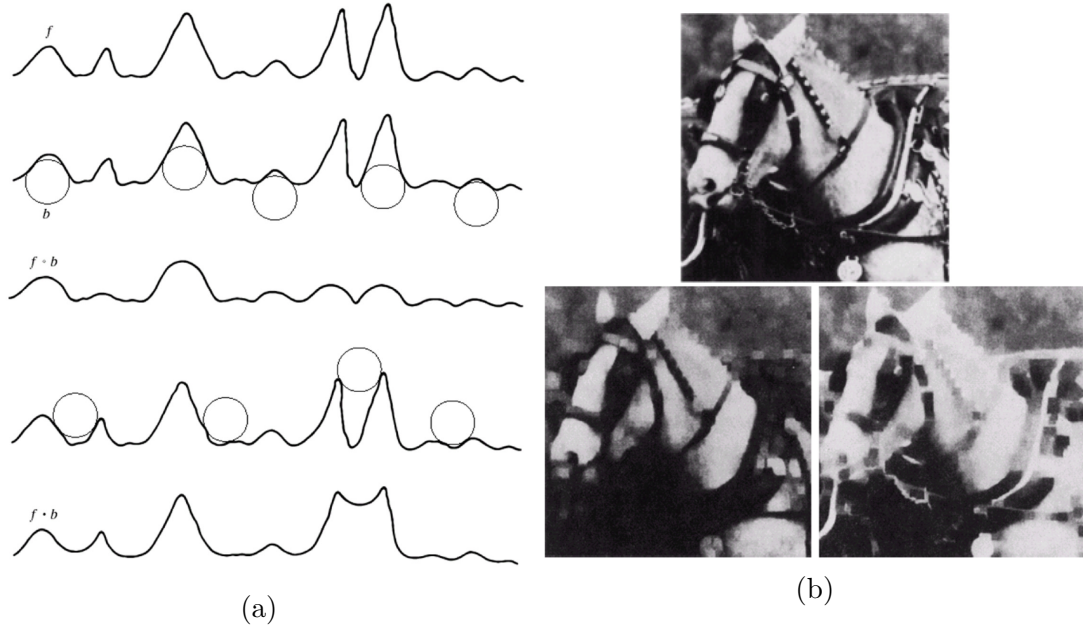


Figure 3.1: The effects of opening and closing on a 1D intensity profile (3.1a) and on a sample image, respectively (3.1b). For the intensity profile, opening eats away the spurious part, whereas closing fills up the valleys between intensity spikes. This behavior can be observed in the image in (3.1b). The original image is given in the upper-right hand corner. The lower-left corner shows the results of a grayscale opening, where small bright details are lost with no great effect on the dark details. The lower-right corner indicates the results of closing, where small dark details are lost with no great effect on the brighter details [7].

### 3.3 Morphological Operations

Mathematical morphology can be described as shape-oriented operations that “simplify image data, preserving their essential shape characteristics and eliminating irrelevancies” [133]. These operations can be explained through set theory. In this context, sets denote shapes in images. For example, sets in 2D Euclidean space represent foreground parts in a binary image, while sets in 3D Euclidean space represent a grayscale image. These operations can be extended and applied to higher dimensional space.

The dilation operation, also known as *expand* or *grow*, is defined for sets  $A$  and

B in space  $E^N$  where a and b are N-tuple points included in these sets as

$$A \oplus B = \{c \in E^N | c = a + b \text{ for some } a \in A \text{ and } b \in B\} . \quad (3.1)$$

From an image processing perspective, B can be considered the structuring element that defines the geometry of expansion. The other fundamental operation is erosion, which is also called *shrink* or *reduce*. The set formula for this operation is

$$A \ominus B = \{x \in E^N | x + b \in A \text{ for } \forall b \in B\} . \quad (3.2)$$

It can be understood from this equation that set B is not contained wholly within A at A's boundary; hence, the boundary pixels of A are not included in set  $A \ominus B$ . In short, the boundaries are etched. Erosion and dilation are usually applied in pairs, and depending on their order, these pairs are called with different names. First one is opening which can be defined as

$$A \circ B = (A \ominus B) \oplus B , \quad (3.3)$$

which can be used to eliminate spurious elements smaller than the structuring element. The second one is closing, which is

$$A \bullet B = (A \oplus B) \ominus A . \quad (3.4)$$

This operation can be used to fill in the gaps smaller than the structuring element, as shown in Fig. 3.1.

These operations set the foundation for a frequently used edge gap filling function in our applications [134]. This function dilates edge ends with a ball structuring element. This is followed by erosion. If two ends that belong to different edges become connected with the dilation, the erosion thins the connection point to the thickness of one pixel. Ultimately, the gap is filled.

### 3.4 Image Denoising in Subcellular Structures

The principal causes of distortion in 3D images are described in [135] as follows: “out-of-focus light, spatial fluctuation of illumination in the specimen, undesired excitation due to lamp flicker, self absorption, bleaching effect and the Poisson noise caused by background photon emission,” all of which result in the 3D impulse response of the microscope (point spread function [PSF]), as seen in Fig. 3.2<sup>1</sup>.

The imperfections of the imaging system can be statistically modeled with two types of noise: intrinsic and extrinsic. Intrinsic noise is caused by the emission of a random number of light photons that occurs when photons from the object reach the detector screen. Random extrinsic noise is triggered by other components of the system during image acquisition. One common reason for this is the charge coupled device (CCD) camera, which is employed as an image detector. Although total noise is best captured with a Poisson distribution, an additive Gaussian noise model can be utilized for computational ease. Denoising approaches can be categorized according to noise model. Among the available methods are non-linear methods and implicit regularization methods, such as Tikhonov regularization, total variation, and Wiener filtering, among others [136, 137]; statistical methods that involve maximum likelihood and maximum a posteriori [135, 138, 139]; and methods utilizing wavelet analysis [140].

Low-rank representations of images and diffusion filters are common practices in medical image denoising [141, 142, 143, 144].

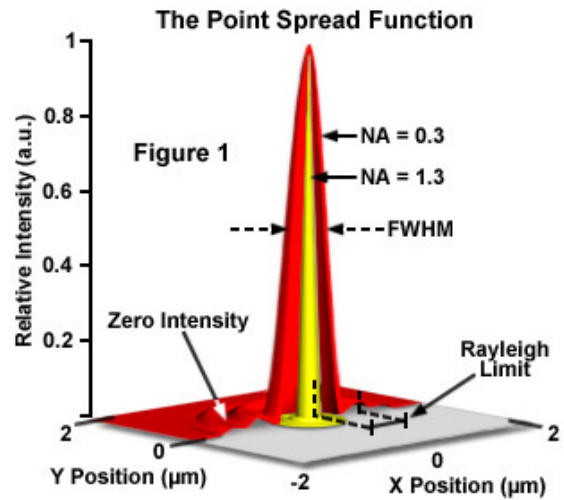


Figure 3.2: Point spread function [8].

<sup>1</sup>Zeiss Education in Microscopy and Digital Imaging



### 3.4.1 Low Rank Image Representation

Low-rank denoising is a strong candidate for use in our initial denoising step. With this method, a return to low-rank subspace might eliminate most of the noise but would sacrifice some of the original signal in the data mixed with the noise (*e.g.* noise).

Given noisy signal  $\tilde{X} = S + X$ , where  $S = \sum_{i=1}^r \theta_i u_i v_i^H$  is lower rank original signal, and  $X$  is random noise-only signal.  $S$  can be recovered under certain conditions through the optimization problem

$$\hat{S} = \arg \min_{\text{rank}(S)=r} \|\tilde{X} - S\|_F, \quad (3.5)$$

whose solution is

$$\hat{S} = \sum_{i=1}^r \hat{\sigma}_i \hat{u}_i \hat{v}_i, \quad (3.6)$$

where  $\|\cdot\|_*$  is Frobenius norm and  $\tilde{X} = \sum_{i=1}^r \hat{\sigma}_i \hat{u}_i \hat{v}_i$  is the singular value decomposition (SVD) of  $\tilde{X}$ . The idea of low-rank representation of a noisy signal is extended in [145]. There, it is argued that the above optimization is a good estimator for solving “the *representation* problem for the best rank approximation of the noisy signal.” However, it is not considered the optimal solution for the denoising problem. Instead of using direct singular value approximation, the researchers devised a weighted approximation. They defined the denoising problem as the squared error between a “weighted low-rank representation” of ( $\tilde{X}$ ), and the recovered signal  $S$  as

$$\text{SE}(w) = \left\| S - \sum_{i=1}^D w_i \hat{u}_i \hat{v}_i^H \right\|_F^2, \quad (3.7)$$

$$w^{opt} := \arg \min_{w=[w_1 \dots w_r]^T \in \mathbb{R}_+^r} \text{SE}(w). \quad (3.8)$$

Note that case  $w_i = \hat{\sigma}_i$  recovers the SVD of  $\tilde{X}$ . The optimization in Eq.(3.8), in terms of weights  $w_i$ , gives the denoised signal through regularizing shrinkage

$$\hat{S}_{opt} = \sum_{i=1}^{\hat{r}} \hat{w}_{i,\hat{r}}^{opt} \hat{u}_i \hat{v}_i. \quad (3.9)$$

### 3.4.2 Diffusion Filters

Physically speaking, diffusion refers to the spread of mass from high-concentration locations to low-concentration locations without any mass being created or destroyed. This physical process can be mathematically modeled by first defining the flux as

$$J = -D \cdot \nabla I , \quad (3.10)$$

where  $D$  is the *diffusion tensor* and  $I$  is the intensity value of image, in the case of image processing [146]. The negative sign on the right-hand side indicates that the diffusion process occurs from areas of high concentration to areas of low concentration. This is because the gradient vector points in the direction of the highest increase so that the diffusion is parallel to the direction of decrease. If  $J$  and  $\nabla I$  are parallel, the case is called *isotropic diffusion* and the diffusion tensor can be given by a scalar *diffusivity*, conventionally called  $g(\cdot)$ . If the function in Eq.(3.10) is a constant or, in other words, independent of position, the diffusion process is referred to as *homogeneous*. If, otherwise, the function depends on the image coordinates, the process is described as *inhomogeneous* [146]. If  $J$  and  $\nabla I$  are not parallel, it is called *anisotropic diffusion*.

As matter is not created or destroyed, the continuity equation can be applied to obtain

$$I_t = -\text{div}(j) = \text{div}(D \cdot \nabla I) , \quad (3.11)$$

where  $I_t$  represents the partial derivative of the image with respect to time, and the negative sign on the right-hand side indicates that the intensity values decrease for positive outward flux values since divergence depicts the net outward flux.

Before designing the diffusion tensor for anisotropic diffusion, it is beneficial to compare linear and non-linear isotropic diffusion. In isotropic diffusion, since the flux and gradient are parallel, the diffusion equation can be given as

$$I_t = \text{div}(g(I)\nabla I) , \quad (3.12)$$

where  $g(\cdot)$  is a scalar function. If it is necessary to preserve the edges of a given initial image  $I$ , the diffusivity is a decreasing function of the gradient magnitude

of the initial image. The pioneering Perona-Malik filter [147] proposes to use the diffusion

$$\frac{\partial I}{\partial t} = \nabla \cdot (g(\|\nabla I\|^2) \nabla I) , \quad (3.13)$$

$$I(t = 0, x, y) = I_0(x, y) , \quad (3.14)$$

with Neumann boundary conditions.

When we perform the operations explicitly we obtain as in [148]

$$\frac{\partial I}{\partial t} = \nabla \cdot (g(\|\nabla I\|^2) \nabla I) , \quad (3.15)$$

$$= \frac{\partial}{\partial x} ((g(\|\nabla I\|^2) I_x) + \frac{\partial}{\partial y} ((g(\|\nabla I\|^2) I_y) , \quad (3.16)$$

$$= g(\|\nabla I\|^2)(I_{xx} + I_{yy}) + 2g'(\|\nabla I\|^2)(I_x^2 I_{xx} + 2I_x I_y I_{xy} + I_y^2 I_{yy}) . \quad (3.17)$$

If we define the unit vector parallel to the gradient as  $\vec{N} = \frac{\nabla I}{|\nabla I|}$  and the normal unit vector as  $\vec{T} \perp \vec{N}$ , we can simplify Eq.(3.17) as

$$\frac{\partial I}{\partial t} = (2\|\nabla I\|^2 g'(\|\nabla I\|^2) + g(\|\nabla I\|^2)) \vec{N}^T \mathbf{H} \vec{N} + g(\|\nabla I\|^2) \vec{T}^T \mathbf{H} \vec{T} , \quad (3.18)$$

where  $\mathbf{H}$  is the Hessian matrix of image  $I$ . The term on the left controls the diffusion in the direction parallel to the gradient vector and the term on the right controls the diffusion normal to the gradient vector.

For the function  $g(\cdot)$ , it was proposed in [147] to use

$$g(s) = \frac{1}{1 + s/K} , \quad (3.19)$$

where  $K$  is a constant. For this definition, Eq.(3.18) becomes

$$\frac{\partial I}{\partial t} = \frac{K(K - \|\nabla I\|^2)}{(K + \|\nabla I\|^2)^2} \vec{N}^T \mathbf{H} \vec{N} + \frac{1}{1 + \|\nabla I\|^2/K} \vec{T}^T \mathbf{H} \vec{T} . \quad (3.20)$$

This filter decreases the diffusion in the direction parallel to the gradient. If  $\|\nabla I\|^2 > K$ , the resulting negative diffusion causes mathematical unstability.

For anisotropic diffusion, if it is desired to enhance the edges, the orthonormal eigenvectors for diffusion tensor  $D$  can be constructed as  $v_1 \parallel \nabla I_\sigma$  and  $v_1 \perp \nabla I_\sigma$ . If the eigenvalues to these eigenvectors are chosen as  $\lambda_1$  and  $\lambda_2$  as

$$\begin{aligned}\lambda_1 &= \alpha , \\ \lambda_2 &= g(|\nabla I_\sigma|^2) ,\end{aligned}$$

where  $\alpha$  is a small positive constant in range  $(0,1)$ ,  $g$  is an decreasing function in range  $(0, 1)$ , and  $I_\sigma$  is the Gaussian smoothed image with variance  $\sigma$ .

The  $I_\sigma$  is a good descriptor for the edges. However, it cannot be used as a descriptor for coherence. Thus, for coherence-enhancing anisotropic diffusion, the diffusion tensor needs to be defined differently [9, 148] (Fig. 3.3). For  $I_\sigma$ , we obtain its tensor product

$$J_0(\nabla I_\sigma) = \nabla I_\sigma \nabla I_\sigma^T . \quad (3.21)$$

The eigenvalues of this matrix are parallel and orthogonal to  $\nabla I_\sigma$ . Thus, the eigenvalues give a representation of the orientation. As it is useful to obtain the orientation of a structure within a region, a componentwise convolution with a Gaussian filter is applied, which means that each element of this matrix is smoothed with the Gaussian envelope fit to neighboring pixel values. Hence, the following measure is defined

$$J_\rho(\nabla I_\sigma) = K_\rho * \nabla I_\sigma \nabla I_\sigma^T , \quad (3.22)$$

where  $K_\rho$  is the Gaussian kernel with  $\sigma$  variance. This is called the *structure tensor* of the image at the particular location. The convolution with the Gaussian kernels enables to find the average orientation in a region. The eigenvalues of the structure tensor  $\mu_1 \geq \mu_2$  can be used as a measure of coherence by defining

$$\kappa = (\mu_1 - \mu_2)^2 . \quad (3.23)$$

Then, we can define the diffusion tensor as the matrix having the same eigenvectors with the structure tensor but inverse eigenvalues. This means that the eigenvector of structure tensor that has the lowest eigenvalue is assigned the highest eigenvalue, namely

$$\lambda_1 = \alpha , \quad (3.24)$$

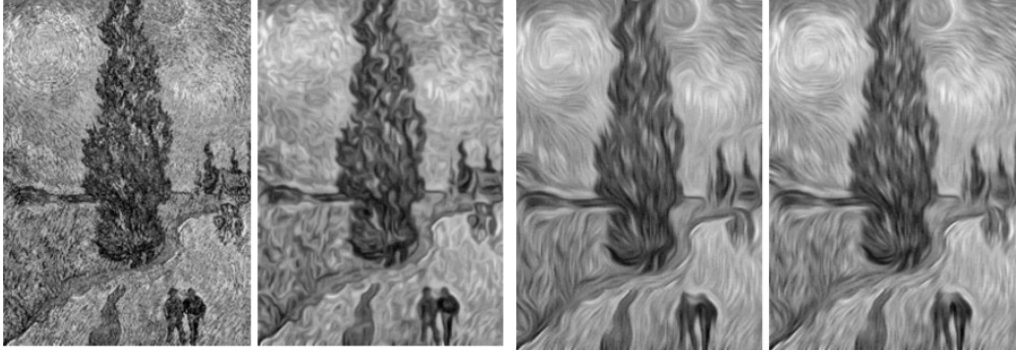


Figure 3.3: Coherence enhancing diffusion applied to Van Gogh’s ‘Road with Cypress and Star’ [9].

$$\lambda_2 = \begin{cases} \alpha, & \text{if } \kappa = 0, \\ \alpha + (1 - \alpha)e^{-C/\kappa}, & \text{otherwise.} \end{cases} \quad (3.25)$$

where  $\kappa$  is the coherence defined in Eq.(3.23). The diffusion tensor can then be found by  $D = Q\Lambda Q^{-1}$ , where  $Q$  is the matrix having eigenvalues as columns.

Another design of the diffusion tensor used for enhancing tubular structures can be found in [141]. The researchers proposed a method of designing the diffusion tensor that uses the eigenvalue analysis proposed by Frangi, which will be detailed in the following section. With the diffusion tensor defined as the diagonal matrix

$$\lambda_1 = 1 + (\omega - 1) \cdot V^{\frac{1}{s}}, \quad (3.26)$$

$$\lambda_2 = \lambda_3 = 1 + (\epsilon - 1) \cdot V^{\frac{1}{s}}, \quad (3.27)$$

where  $\lambda_1$  is the eigenvalue corresponding to the eigenvector associated with the smallest eigenvalue of the Hessian matrix. In the equations,  $V \in [0, 1]$  and  $\omega > \epsilon$ ,  $\epsilon > 0$  and  $s$  is a positive real constant. The parameter  $\omega$  denotes the strength of anisotropic diffusion,  $\epsilon$  should be a small value to represent the isotropic diffusion component, and  $s$  is the term used to control the weight of the vesselness.

With this definition of the diffusion tensor, diffusion can be performed along the tubular structure, in the direction of the eigenvector corresponding to the smallest eigenvalue of the Hessian matrix of the image.

## 3.5 Tubular Structure Segmentation

Computer vision research has been aiming to help medical doctors and biologists by automating the detection of tubular structures. Different imaging modalities exist for different detection problems to capture the structures of interest in the best way possible. In different fields, 3D and 2D imaging data are utilized for various purposes in analyzing tubular structures. Computer tomography (CT), positron emission tomography (PET), magnetic resonance, ultrasound are commonly used imaging modalities for analysis. For instance, retinal scans, and angiographs in medical field are commonly obtained in 2D for vasculature [149]. On the other hand, maximum 2D projections of confocal microscopy images of MTs are utilized for colocalization demonstrations, although the analysis was done in 3D by [48].

As there are different imaging modalities for different research purposes, there have been many different methods trying to solve the similar problems. Some of these methods have been based on models or fixed filters. Some studies approached the subject from active contours perspective due to their topological advances. Machine learning stepped in when the researchers had large enough labeled data. Recent developments in neural networks research also have been employed to tackle the problem.

### 3.5.1 Filter/Model Based Methods

These methods apply designed filters searching for meaningful results or look for structures that fit a certain mathematical model.

#### Hessian Based Methods

One of the predominant studies on vessel segmentation is [40]. The researchers described the importance of knowing the vessel structure for different medical

applications. The problems faced during the acclimation of the vessel structure are the overlap of non-vascular structures and the low visibility of small vessels.

According to the researchers, vessel enhancement is a filtering process that searches for geometrical structures (tubular structures in our case). Taylor expansion is used to analyze and observe the local behavior of an image in the neighborhood of a point up to the second order.

$$L(x_o + \delta x_{o,s}) \approx L(x_o, s) + \delta x_o^T \nabla_{o,s} \mathcal{H}_{o,s} \delta x_o . \quad (3.28)$$

In Eq.(3.28), multipliers of the last term express the gradient vector and Hessian matrix of the image calculated in  $x_o$  with scale  $s$  or, in other words, the derivative of the signal. To smoothly compute these differential operators of  $L$ ,  $L$  was convolved with the derivatives of Gaussians (DoG).

$$\frac{\delta}{\delta x} L(x, s) = s^\gamma L(x) * \frac{\delta}{\delta x} G(x, s) . \quad (3.29)$$

$\gamma$  in this equation is used to describe a family of normalized derivatives. It is important when we compare the response of differential operators at multiple scales.  $\gamma = 1$  for no scale.

The second derivative of a Gaussian kernel at scale  $s$  provides a probe kernel that measures the contrast between the regions inside and outside the range  $(s, -s)$  in the direction of the derivative. Hessian gives us a matrix of directional derivatives.

$$\delta x_o^T \mathcal{H}_{i,j} \delta x_o = \left( \frac{\partial}{\partial \delta x_o} \right) \left( \frac{\partial}{\partial \delta x_o} \right) L(x_o, s) . \quad (3.30)$$

With the eigenvalue analysis of Hessian, we obtain the principal directions in which the local second structure of the image can be decomposed. For the ordering of eigenvalues as  $|\lambda_1| \leq |\lambda_2| \leq |\lambda_3|$ , a pixel belonging to a vessel region will have  $|\lambda_1| \approx 0$  since the second order characteristics will have small change in the direction of the vessel and  $|\lambda_1| \ll |\lambda_2| \approx |\lambda_3|$ , when  $\lambda_2, \lambda_3$  are negative since the vessels show as light tubular structures on dark background in CT scans (Table 3.1). According to the values of  $\lambda$ 's, we can describe what kind of region the pixel is in as blob,

$\lambda_1$	$\lambda_2$	$\lambda_3$	structure
Low	Low	Low	noise
Low	Low	High-	plate-like (bright)
Low	High-	High-	tubular (bright)
High-	High-	High-	blob-like (bright)

Table 3.1: The structure analysis is based on the eigenvalues of the Hessian matrix when the eigenvalues are ordered as  $|\lambda_1| \leq |\lambda_2| \leq |\lambda_3|$ , where a positive sign indicates positiveness. This classification applies to bright objects against a dark background. For dark objects on a bright background, the signs should be reversed.

plate or tube. In [40], three ratios are defined as,

$$R_A = \frac{|\lambda_2|}{|\lambda_3|}, \quad (3.31)$$

$$R_B = \frac{|\lambda_1|}{\sqrt{|\lambda_2\lambda_3|}}, \quad (3.32)$$

$$S = \sqrt{\sum_{i \leq D} \lambda_j^2}, \quad (3.33)$$

where  $D$  refers to the dimensions of the image. The ratio  $R_A$  compares the eigenvalues in the directions perpendicular to the least change. The closer this value is to 1, the more the voxel is a part of a plate-like structure. The closer it is to 0, the more the voxel is a part of line-like structure. The ratio  $R_B$  indicates whether the voxel is a part of blob-like or a plate-like structure. For blob-like structures, the eigenvalues are expected to be close to each other since the intensity change in each direction is similar in magnitude. Thus, the ratio approaches 1 for blob-like structures. The parameter  $S$  is the Euclidean norm of the eigenvalues and is called the second-order structureness. This parameter works to eliminate the noise in the image, while accounting for the strength of the second-order characteristics of the image.

As discussed in [40], a vessel, because of its circular cross-section, has a plate-like characteristic perpendicular to the vessel axis and the vessel is not blob-like, as shown in Fig. 3.4. Using these two characteristics of the model, the researchers suggested a vesselness measure using the parameters described above.




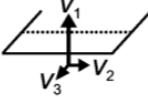
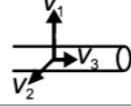
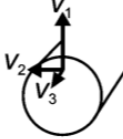
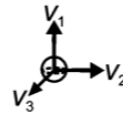
	$\mu_1 \approx \mu_2 \approx \mu_3 \approx 0$
	$\mu_1 \gg \mu_2 \approx \mu_3$
	$\mu_1 \approx \mu_2 \gg \mu_3$
	$\mu_1 > \mu_2 > \mu_3$
	$\mu_1 \approx \mu_2 \approx \mu_3 \gg 0$

Figure 3.4: Illustration of the eigenvectors  $V_i$  (Column 1) and corresponding eigenvalues  $\mu_i$  (Column 2) of the structure tensor in five different situations. The first case indicates almost no contrast in any direction; the eigenvalues are equal and close to 0. The second case shows contrast in the direction of  $V_1$ ; the first eigenvalue is much larger than the other two eigenvalues. The third case shows contrast in the direction of  $V_1$  and  $V_2$ ; the first two eigenvalues are much larger than the third. In the fourth case, there is great contrast in  $V_1$ , less contrast in the direction of  $V_2$ , and almost no contrast in the direction of  $V_3$ . The fifth case indicates contrast in all directions; the eigenvalues are equal but are much larger than 0.

The measure follows as

$$V_o(s) = \begin{cases} 0, & \text{if } \lambda_2 > 0 \text{ or } \lambda_1 > 0, \\ (1 - \exp(-\frac{R_A^2}{2\alpha^2})) \exp(-\frac{R_B^2}{2\beta^2}) (1 - \exp(-\frac{S^2}{2c^2})), & \text{otherwise.} \end{cases} \quad (3.34)$$

where  $\alpha$ ,  $\beta$ , and  $c$  are constants that control the sensitivity of the parameters.  $V_o(s)$  approaches 1 as the structure becomes tubular.

### COSFIRE Filter for Vessel Delineation

In this method, a COSFIRE (Combination of Shifted Filter Responses) filter is employed to obtain a high response in tubular structures [42]. The intensity maxima at points are aligned on concentric circles. The values at these points are

represented as a set of tuples  $S$  of DoG sigma,  $\sigma$ , and polar coordinates  $\rho$  and  $\phi$ . The filter response is computed as the geometric mean of the tuples set after the DoG images are blurred by a Gaussian function where  $\sigma' = \sigma'_0 + \alpha\rho_i$  and are shifted in the opposite direction of tuple  $\phi_i$  with amounts  $\delta x_i = -\rho_i \cos \phi_i$  and  $\delta y_i = -\rho_i \sin \phi_i$ .

$$r_s(x, y) \stackrel{\text{def}}{=} \left| \left( \prod_{i=1}^{|S|} (s_{\sigma_i, \rho_i, \phi_i}(x, y)) \right)^{1/\sum_{i=1}^{|S|} \omega_i} \right|_t. \quad (3.35)$$

### Wavelet Approach

Wavelets are wave-like oscillations where the amplitude starts at 0, increases, and returns to 0. Wavelets can be convoluted with an unknown signal to obtain information. In image processing, wavelets are employed to decompose images into different levels of details according to their scaling and frequency. Symlet wavelets are used to decompose images [41]. These are compactly supported wavelets with the least asymmetry (adhering to the vessel structure) and the highest number of vanishing moments for a given support width (capturing contrast changes on the sides of a tubule). Their associated scaling filters are near linear-phase filters.

### A Combination Morphological Operations and Differential Image

In one of the methods, the grayscale images are first morphologically processed with one-pixel-wide operators to obtain an initial idea of possible vessel areas. At first, linear structures are accentuated by taking the supremum of image opened with different orientation filters [43]. In addition, non-structure regions are obtained through the infimum of images filtered with orthogonal directions. The output is passed through a Laplacian of Gaussian (LoG) filter in different orientations. Another morphological operation is used to remove the nonlinear brightness regions where there are no tubular structures. In the case of morphology filters, the size of the morphology filter and sigma of LoG are the determining factors of segmentation quality.

## Differential Operation Based on Kirsch Templates

In this approach, differential filter templates such as  $\begin{bmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{bmatrix}/15$ ,  $\begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix}/15$ ,  $\begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix}/15$ ,  $\begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix}/15$ ,  $\begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ -3 & 5 & 5 \end{bmatrix}/15$ ,  $\begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}/15$  are convolved with the image. If the filter structure is observed, it is seen that these filters are searching for certain intensity change profile similar to change profile for Frangi filter. For each response image, the maximum response is chosen and included in output image. If pixel values are above a given threshold, they are considered as tubular structures [150].

### 3.5.2 Energy Minimization

These methods are usually based on manipulating images or segmenting structures according to the optimization of an energy function.

#### Geodesic Active Contours

This approach improves the classical method that depends on ‘deforming’ an initial contour  $C_0$  towards the boundary of the object of interest. This method utilizes the minimization of a functional so that we get the boundary at its local minimum [151]. This functional is composed of two parts, one that is controlling the smoothness, and the other attracting the curve towards the boundary. These two parts can be thought of the velocity parts of the shrinking or expanding curve. Again the first one is related to the curve’s regularity and the second makes the curve shrink or expand. The main mathematical pillar of this flowing motion stands on the partial differential equation solutions. In [151], the old energy minimization approach of employing “snakes” is adopted for curve evolution. In the energy minimization approach

$$E(C) = \alpha \int_0^1 |C'(q)|^2 dq + \beta \int_0^1 |C''(q)|_2 - \lambda \int_0^1 |\nabla I(C(q))| dq . \quad (3.36)$$

the first two terms are the regularization components and the last term is the shrinking component. The new method eliminates the term with  $\beta$  coefficient.

The functional thus becomes

$$E(C) = \alpha \int_0^1 |C'(q)|^2 dq - \lambda \int_0^1 |\nabla I(C(q))| dq . \quad (3.37)$$

The regularization is obtained through the consideration of the curvature in geodesic curves. We find the curve at the points of maxima  $|\nabla I|$  (i.e., an edge detector) while minimizing this functional. A generalization is obtained by replacing  $|\nabla I|$  with  $g(|\nabla I|)^2$ , where  $g(r)$  is a strictly decreasing function.

A relation between energy minimization and geodesic curves is obtained through Maupertuis' principle. This theorem states that curves that are extremal in Euclidean space and have fixed energy  $E_0$  are geodesic curves with non-natural parameters. These curves have geodesic property with respect to a new metric  $g_{ij} = 2m(E_0 - U(C))$ , where  $U(C) = -\lambda g(|\nabla I(C)|)^2$ . Hence, minimizing  $E(C)$  would yield

$$\int_0^1 \sqrt{g_{ij} C'_i C'_j} dq ,$$

$$\int_0^1 \sqrt{g_{11} C_1'^2 + 2g_{12} C_1' C_2' + g_{22} C_2'^2} \quad i, j = 1, 2 .$$

Using Fermat's principle of light, we can reduce the minimization problem to  $\int_0^1 g(|\nabla I(C(q))|) |C'(q)| dq$ . We minimize this expression by calculating the gradient descent direction. Thus, we solve Euler-Lagrange equation.

$$\frac{\partial C(t)}{\partial t} = g(I) \kappa \vec{N} - (\nabla g \cdot \vec{N}) \vec{N} .$$

We can express  $C$  implicitly if we assume that it is a level set of function  $u$ . Then the equation above becomes

$$\frac{\partial u(t)}{\partial t} = g(I) |\nabla u| \kappa + (\nabla g(I) \cdot \nabla u) \vec{N} . \quad (3.38)$$

Surface extraction through the level set method can be utilized for refining vessel segmentation after the vesselness measure is found. However, this method heavily depends on parametrization, which makes it difficult to apply.

Another method proposed in [152] is to apply minimization on the image, taking the vesselness maps obtained at different scales and using the method described in the previous section. Each voxel of the image contains a vector that consists of the vesselness values obtained in different scales. For the initial and final points, the algorithm starts from each vesselness value in the initial voxel to reach one of the vesselness values of the final voxel by passing through the neighboring pixels and scales. The cost function takes into account the intensity difference between the neighboring pixels and the jump in scale. By applying this definition of the cost function at different scales, the minimizing path does not contain sudden jumps in scales. The algorithm accordingly prevents kissing or crossing vessels.

### Stretching Active Contours

In [153], stretching open active contours (SOACs or SOAX) are formulated as open-ended parametric curves with stretching open ends. The curves extend or shorten to fit structures. As is the case with many active contour-based methods, the curves evolve to minimize an energy functional, which is composed of internal energy and external energy. The internal energy component ensures the smoothness and continuity of the curve  $\mathbf{r}(s)$ . On the other hand, external energy makes sure the curve follows image components of interest such as edges and ridges, among others. In the SOAC case, external energy adds stretch energy on top of image-related potential energy [46].

$$\mathcal{E}_{int}(\mathbf{r}) = \int_0^L \alpha(s) |\mathbf{r}'(s)|^2 + \beta(s) |\mathbf{r}''(s)|^2 ds , \quad (3.39)$$

$$\mathcal{E}_{ext}(\mathbf{r}) = \int_0^L k_{img} E_{img}(\mathbf{r}(s)) + k_{str} E_{str}(\mathbf{r}(s)) ds , \quad (3.40)$$

where  $\alpha$  and  $\beta$  are weights for adjusting curve tension and smoothness and  $k_{img}$ ,  $k_{str}$  are weights for image and stretching forces.

The image force can be formulated as  $\nabla(I * G_\sigma)(\mathbf{r}(s)) = (I * \nabla G_\sigma)(\mathbf{r}(s))$ . The stretch magnitude is computed as

$$F(\mathbf{r}(s)) = \frac{I(\mathbf{r}(s)) - I_{lb}(\mathbf{r}(s))}{I(\mathbf{r}(s))}, \quad s = 0, L, \quad (3.41)$$

where  $I(\mathbf{r}(s))|_{s=0,L}$  is the image intensity at the curve's tips and  $I_{lb}(\mathbf{r}(s))|_{s=0,L}$  is the intensity of the local background around each tip [46]. The tangential stretching force is designed as

$$\mathbf{F}(\mathbf{r}(s)) = \begin{cases} -\frac{\mathbf{r}'(s)}{|\mathbf{r}'(s)|} \cdot F(\mathbf{r}(s)), & \text{if } s = 0, \\ \frac{\mathbf{r}'(s)}{|\mathbf{r}'(s)|} \cdot F(\mathbf{r}(s)), & \text{if } s = L, \\ 0, & \text{if } 0 < s < L. \end{cases} \quad (3.42)$$

Combining two force fields together, the external force is

$$\nabla \mathcal{E}(\mathbf{r}(s)) = k_{img} \cdot (I * \nabla G_\sigma)(\mathbf{r}(s)) + k_{str} \cdot \mathbf{F}(\mathbf{r}(s)). \quad (3.43)$$

In [46], the framework for a single SOAC is extended to the sequential evolution of multiple snakes and achieved through snake overlap checks during and after the evolution.

### 3.5.3 Learning Based Methods

Some tubule segmentation methods use machine learning-based approaches to segment vessels in order to leverage the high amount of visual data available. Some approaches utilize Markov/Conditional random fields [154, 155, 156]. Traditional feature engineering-based learning has also been employed in some studies [157, 158, 159, 160]. With the resurgence of neural network research, researchers have developed methods applying deep learning or a combination of deep learning with other representation methods to tubular structure detection [161, 162, 163, 164].

### 3.6 Perceptual Grouping-Tensor Voting

Tensor voting is a method that has been proposed to perceptually group and extract visually salient elements in spite of gaps in features and noise [10]. Tensor voting takes place in four stages. The data is first encoded into the second-order symmetric tensor  $T$ , which can represent certainty and uncertainty in the direction of visual structures as a combination of stick and ball tensors, respectively. This combination can be formulated as

$$T = (\lambda_1 - \lambda_2)\hat{e}_1\hat{e}_1^T + \lambda_2(\hat{e}_1\hat{e}_1^T + \hat{e}_2\hat{e}_2^T), \quad (3.44)$$

where  $\lambda_1$  and  $\lambda_2$  are eigenvalues of  $T$ , and  $\hat{e}_1, \hat{e}_2$  are associated eigenvectors. The term  $(\lambda_1 - \lambda_2)$  is called the *saliency* of the tensor, and it measures confidence in the encoded direction of the visual structures.  $\lambda_2$  is called *junctionness*, and it represents the level of contradiction in the encoded direction of the structures. Votes cast by this scheme can also be represented using tensors. The fundamental voting tensor is called a stick voting field. It is designed to cast a vote by using the parameters of the osculating circle connecting the voting point and the vote location, as shown in Fig. 3.5. The strength of the vote is enveloped with a decay function that punishes distance and high curvature.

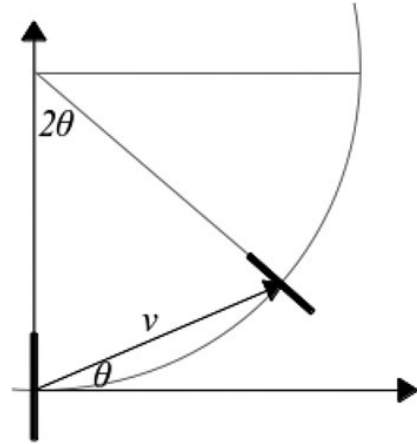


Figure 3.5: The osculating circle connecting the voting point and voting location [10].

$$DF(\mathbf{v}) = e^{s^2 + c\kappa^2/\sigma^2}, \quad c = \frac{-16 \ln(0.1) \times (\sigma - 1)}{\pi^2}, \quad (3.45)$$

where  $s$  is the arc length along the osculating circle,  $\kappa$  is the curvature,  $\sigma$  is the neighborhood size, and  $c$  is the decay control variable. The voting tensor is rotated and scaled by the decay to compute the vote. Votes are collected through tensor addition. The resulting tensor points to the structures that seem to adhere to

perceptual continuity. Fig. 3.6 shows the propagation of the voting.

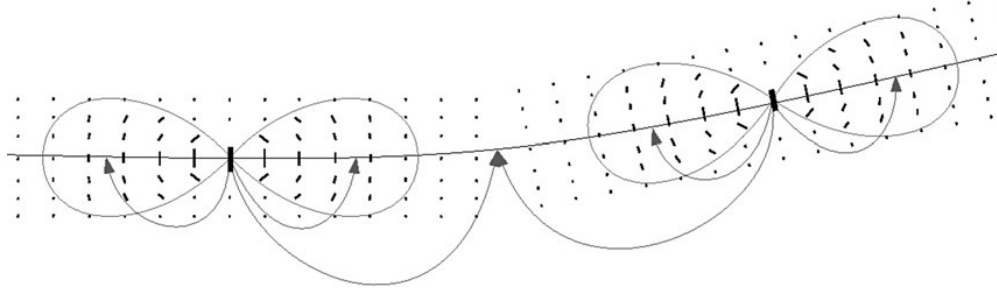


Figure 3.6: The stick voting field and its propagation [10].

### 3.7 Summary

As seen above, exhaustive research in tubular structure detection and denoising has been conducted. When these approaches are evaluated in light of our stoma microtubule detection problem, a few challenges arise. Although active contour-based methods are useful for addressing topological changes, the initial structure may influence results. The continuity breaks when the external force (image) has problems. Filter-based methods are robust in detecting the shapes that fit their mathematical model and may be useful starting points as an input to other methods. However, they are also susceptible to artifacts and brittle to structures deviating from their inherent model. On the other hand, deep learning models require an exhaustive amount of data, which we currently do not have.



# **Beyond Saliency: Assessing Visual Balance with High-level Cues**

## **4.1 Introduction**

In this chapter, we present the details of our study of visual saliency representing compositional balance. As the background provided in Chapter 2 indicates, saliency fails to capture some elements that help in creating a balanced composition. We collect an image dataset along with visual center ratings. We show saliency is not a good representation of an image’s visual center and propose another representation that uses additional high-level visual features.

The chapter starts with the motivation regarding the problem with employing saliency for visual center representation. After pointing out the problem, we describe our approach to visual center representation. The data collection process, relevant feature extraction, and validation steps are discussed. Finally, we give a discussion about the limitations of our approach.

### **Motivation**

We expand on the saliency map concept that is frequently employed in the aforementioned computational aesthetic studies. We also show the problem by presenting cases in which bottom-up saliency methods fail in visual balance representation. As explained in Chapter 2, bottom-up approaches consider the pre-attentive vi-

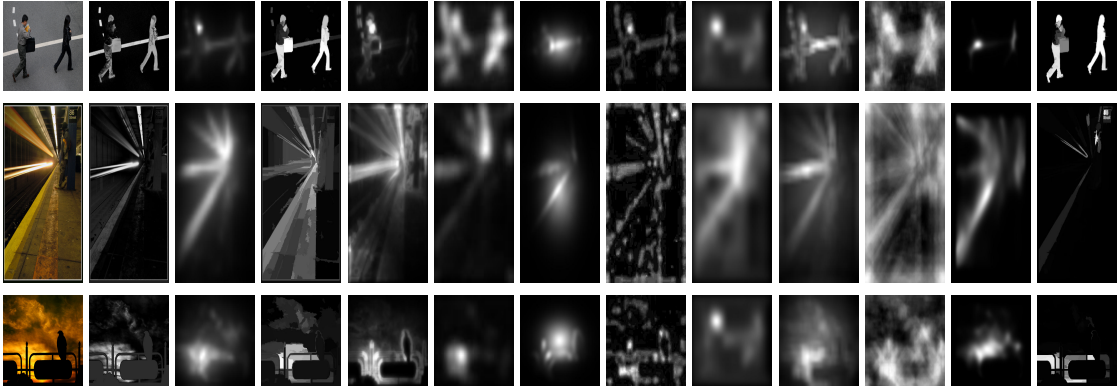


Figure 4.1: Saliency analyses often fail to capture the key elements in attention. Columns: original image, Achanta, GBVS, DRFI, Context aware, imSig, covSal, corSal, UHF, SWD, Murray, FES, MDF.

sual attention that plays a role in a viewer’s unconscious reaction. This nature constrains these methods more to low-level features that are processed in earlier stages of the human visual system. When the center of mass (CoM) of a bottom-up saliency map is evaluated to determine whether visual attention is lopsided or aligned with the physical center of an image, some problems arise. Fig. 4.1 shows a few examples where saliency methods fail to capture the elements that actually have more or less visual weight than the saliency map assigns to them. The first picture in Fig. 4.1 contains two humans that balance each other visually. At first glance, almost half of the methods tested fail to capture the humans, even though it was shown in a study that humans, faces, body parts, text bodies, and animals receive more attention than other objects [79]. Among the methods tested, the ones that successfully capture humans in pictures assign attention to other parts of images that actually do not have that much visual weight compared to humans. The text area above the person in the second image is also mostly missed, but it may not have that much visual weight as it covers a small area in the image. More interestingly, the saliency maps employed capture the light beams of the subway and the lines created by subway station floor to a great extent. It is worth noting that these lines are parallel in reality but appear to converge on the photograph plane. This structure leads the eye from the left to the convergence point, which has a different visual weight. Saliency methods do not succeed in reflecting this concept. The third image shows another failure of a saliency map in images that

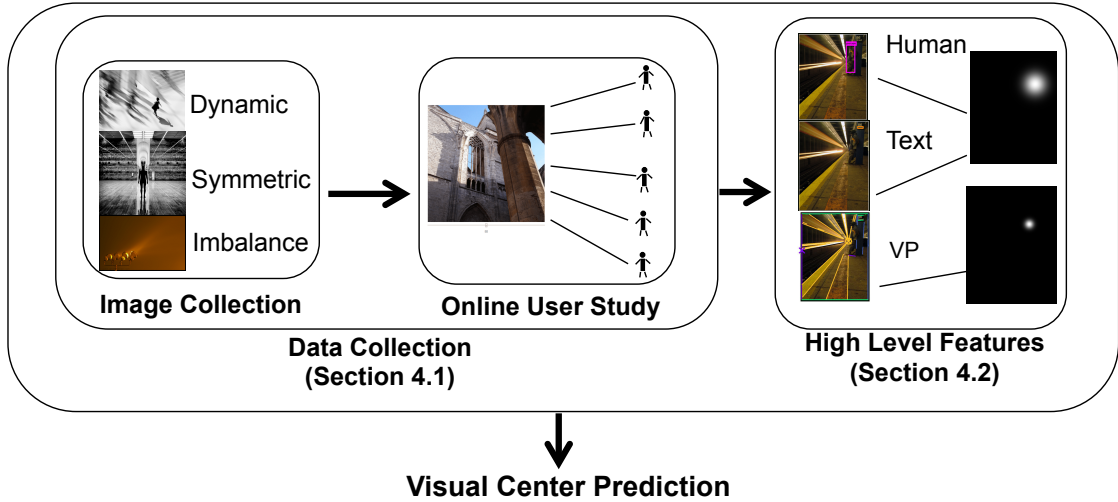


Figure 4.2: Our framework for visual center prediction.

contain high-contrast regions. The bird silhouette perched on the railing is surrounded by a relatively light region of clouds. Saliency methods successfully give the outline of the bird and the railing bars, which are regions of high contrast. Yet in a higher scale the bird itself is different from the cloud structure, so it has more visual weight. These imperfections in saliency maps in terms of visual weight indicate that a new approach that takes high-level features into account could be highly beneficial.

## 4.2 The Method

Our visual center prediction framework contains four parts as summarized in (Fig. 4.2). The first two parts are related to obtaining an image set and participant ratings related to the visual centers of the images. The third component is the extraction of low- and high-level features and the translation of these features into a vector representation. Finally, we predict the visual center of each image through a linear regression scheme. This section provides details about the first three components.

### 4.2.1 Data Collection

We detail the image set creation, interface design for acquiring ground-truth labels, and feature extraction from the image set.

**Image Set.** The visual data for the initial study was mainly collected from a popular photo-sharing Website, Flickr using the query “urban.” The images were ranked according to their interestingness. They were categorized into four classes– “symmetrical balance,” “dynamic balance,” “imbalanced,” and “hard to tell”–by the authors based on consensus. While categorizing the images, the definitions in [4, 18] were taken into consideration. After eliminating the images that were not related to the concept of balance, 779 images remained. In the end, the number of samples in each class were 41, 362, 90, and 447, respectively. The fact that the total number of balanced images was 403 and the number of imbalanced images was 90 caused an imbalance in the data. Hence, 286 more imbalanced images were selected based on consensus using the query results for “travel,” “Hawaii,” “jungle,” “party,” “galaxy,” and “vacation.” Hence we had 403 balanced and 376 imbalanced images.

The reason these query words were selected is that urban pictures were dominated by man-made structures. By adding images belonging to these query words we ensured that more natural scenic images with different properties were included in our dataset.

**Interface Design.** An online study was designed to collect data from the participants. The challenges of designing an online survey about compositional balance started with determining an appropriate quantification or measure of balance. In the empirical art field, some studies have involved putting a fulcrum under a given picture and asking the participants to move the lever till the picture feels balanced on the fulcrum [3, 23]. In these studies, the

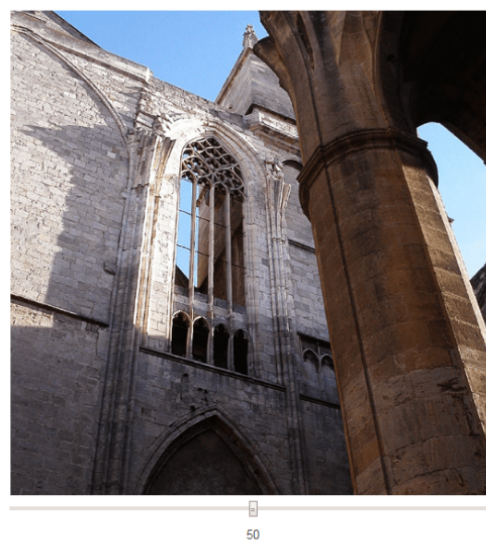


Figure 4.3: User interface for online human subject study. The slider bar below the image allows the participant to indicate the visual center.

normalized balance scores were recorded in an interval  $[-1, 1]$ , where 0 was the exact center of an image, and -1 and 1 were the extreme edges. In a similar fashion, we recorded the balance values in the interval  $[0, 100]$ , where 50 was the exact center of the image and 0 and 100 represented the extreme edges (Fig. 4.3). This is one of the first online balance studies to the best of our knowledge. The second challenge was coming up with a concise yet meaningful tutorial that included a definition of compositional balance and how it is achieved. As mentioned in the previous section, simple definitions of compositional balance and its methods were given in [4, 18], along with sample images. Once the participant taking our study selects a participant ID, he/she must read through a tutorial explaining all of the concepts that are being studied in the survey. The ways compositional balance are achieved are described using sample images and text in the tutorial, which align with the rules that we used to select the images for our image set. After the tutorial, the participant enters the survey. The survey contains a white background with the test image in the center-left and a slider bar beneath the image. The slider bar starts at the mark in the center, which correlates to 50 in our study. The participant is asked to “move the slider in a way that the image feels visually balanced at that point.” If the participant needs any assistance or would like to refer to the tutorial, he/she is free to do so by going back “home.” For each image, we limited the number of ratings to five. Each participant was only presented with a specific image once. In addition, to encourage breaks, each participant completed a batch of only 100 images at once. When the participant had completed the ratings for 100 batches, or there were no more images to rate, he/she entered a screen informing him/her of that. It was possible for the participant to start a new batch right after one was completed. If the participant was inactive on a single page for more than six minutes, the survey timed out and the participant was returned to the home page. By the end of the survey time period, we had collected important data including participant IDs and ratings for the center of mass. The study was taken by eight participants, providing a total of 3895 ratings.

**Human Subject Study and Data Properties.** The ratings were collected from undergraduate, graduate students, and faculty who participated in the study. Each image received five ratings. The average rating position of the slider,  $\bar{p}_i$ , for each image  $i$  was calculated. The mean slider position across all images  $p^*$  was 50.17

with a standard deviation of 6.7. The equation is

$$\bar{p}_i = \frac{1}{N} \sum_{u \in \text{Participants}_i} p_i^u, \quad p^* = \frac{1}{M} \sum_i \bar{p}_i, \quad (4.1)$$

where  $\text{Participants}_i$  is the set of participants who have rated image  $i$ . When the images with the highest standard deviation in the slider position were inspected, border cases in which people showed disagreement were observed. It was understood that the participants may have had different opinions of symmetry and were confused by dynamic balance in some cases. It was also observed that some participants agreed that an image was imbalanced, but their slider ratings pointed to different sides for imbalance. This seemingly suggested that individuals had different understandings of the concept. These results showed that there were cases of agreement among the participants in relation to the images, providing hope for achieving a consensus.

The images where participants had a high level of agreement were chosen using a standard deviation threshold for the slider positions. The standard deviation of the slider position rates was expected to be low where the participants agreed. The thresholding process resulted in a total of 593 images with a high level of agreement. Fig. 4.4 displays some pictures with low (first row) and high (second row) standard deviation. Looking at the high standard deviation images, it is clear that participants assigned different weights to different parts of the same picture. While some participants gave more weight to the man sitting closer to the camera or the girl in the red dress, some deemed the weight distribution equal, as shown in Fig. 4.4d and 4.4f. The weight that the participants assigned to the dark phone booth varied, as seen in Fig. 4.4e.

### 4.2.2 High- and Low-Level Features for Balance

This subsection explores the features that may be helpful in predicting the visual center of an image.

**Saliency.** To model visual attention distribution over an image, saliency maps were employed. Bottom-up saliency approaches were better-suited to our purpose than top-down approaches, because we are interested in unconscious visual

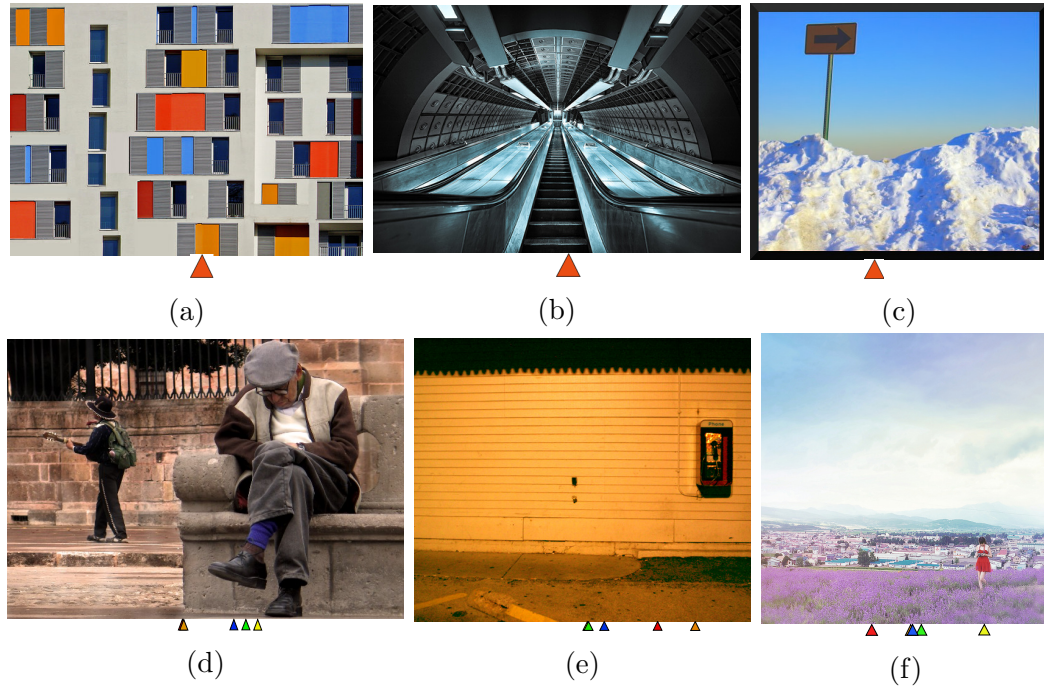


Figure 4.4: Sample images with low and high standard deviations for the average slider position. Images with lower standard deviations are shown in the first row, where slider positions overlap and the position is illustrated with a single triangle. The second row shows images with higher standard deviations. The participant ratings are marked using triangles to show the spread of the slider positions.

responses to composition balance, which depends on low-level features. A representative of each category of saliency methods in Table 2.1 was selected for our study. These methods included context-aware saliency, image signature, graph-based, UHF, MDF, and non-parametric low-level vision saliency (Murray).

**Informative Objects.** People pay more attention to objects that provide different kinds of information about a given picture. These objects can be considered high-level features. The information can include emotional cues, actual knowledge, or the message. One of the prominent sources of information in pictures is humans. Another object class is text in the image. The inclusion of the spatial distribution of these objects along with saliency could improve the representation accuracy of visual balance. Out of the 779 images in our dataset, there were 228 images that contained humans. There were 161 images that contained text regions.

For human detection, a state-of-the-art scene annotation system based on deep neural networks, YOLO, was utilized [165]. The model was trained on ImageNet 1000-class competition data [166]. In this approach, a single convolutional network is run on the whole image. The network returns bounding boxes of objects detected with a probabilistic confidence. Finally, the system thresholds the bounding boxes by their corresponding probabilities. The method employed was successful in returning human positions for 226 out of 228 images. The recall of this method on our dataset was 0.99 in terms of detecting humans. For text detection, another deep learning based method, known as the connectionist text proposal method and proposed in [167], was employed. This approach detects text lines by densely sliding a small window (3x3) in the convolutional feature maps and producing fine-scale text proposals. Fine-scale text proposals are susceptible to the false detection of windows and bricks, among other objects. The sequential structure of text is exploited via final recurrent neural network to improve performance. The recall rate of text detection on our dataset was 0.89 . The detection boxes were eliminated by considering the aspect ratio, as very thin text or humans/human parts are normally not recognized by people.

**Eye Leading Lines.** As previously mentioned, one of the elements that is used to strike compositional balance is eye leading lines (Fig.4.5). The concept of eye leading lines in visual art or photography is related to perspective, which is mainly associated with prominent geometrical structures called *parallel lines* [168].

On a photograph plane, parallel lines appear to converge towards a point, which is called the vanishing point (VP). There may be more than one VP in a scene, but the main focus is on the dominant one. In our study, we leveraged the dominant vanishing point detection method proposed in [168] to account for eye leading lines information, which can be seen as a high-level feature. In this method, straight edges are extracted from ultrametric contour maps by subdividing the contour into straight line segments at points that have maximum distance to the straight line connecting

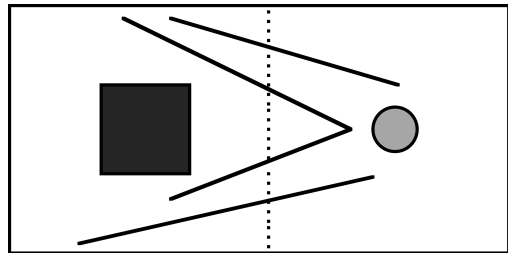


Figure 4.5: Eye leading lines are utilized to strike a balanced composition[4].



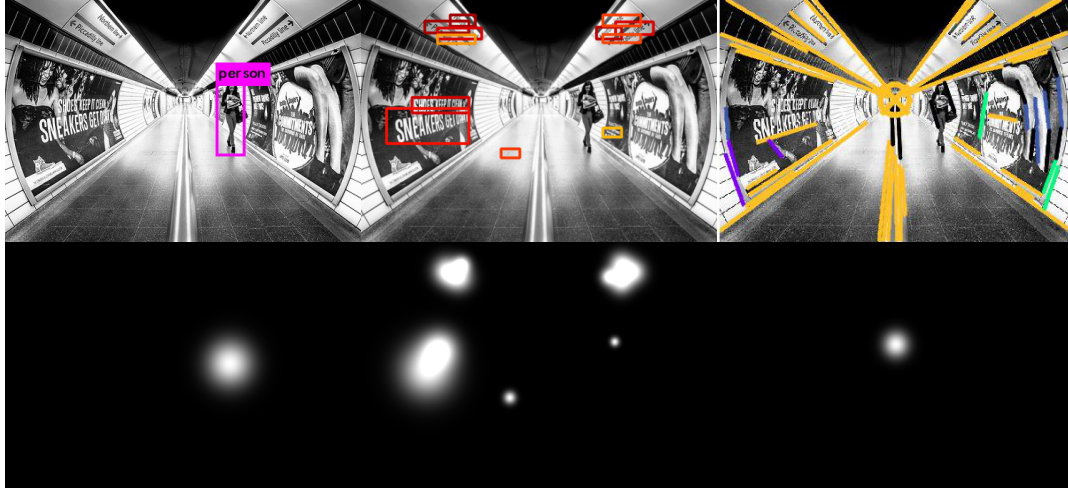


Figure 4.6: Representation of different features. Gaussians are fitted within the boxes obtained and the CoM of this image is calculated. From top-left to bottom-right: YOLO human detection, text detection, dominant VP detection, and the corresponding Gaussian envelopes.

the end points of the contour. Then these straight edges are grouped according to J-Linkage fitting. Two random edges ( $E_{j1}, E_{j2}$ ) are sampled from the edge set. The lines are fitted to each edge and their hypothetical intersection point  $v_j$  is obtained. The J-Linkage creates a preference matrix according to a consistency measure. The consistency measure is defined as the root mean square of the distance between the edge points and line  $\hat{l}$  that passes through the hypothetical  $v_j$  and minimizes this distance. Once the preference set for each edge is acquired, edges that have similar preference sets are clustered together. The  $v_j$  of the biggest cluster becomes the dominant VP.

### 4.2.3 Representation

Current literature in psychology and computer science borrows the physical concept of center-of-mass (CoM) given in Eq.(4.2). In this case, the visual weight elements were modeled as point-masses. The CoM coordinates give an idea of the quadrant into which the visual weight of the image falls.

$$CoM = \frac{1}{M} \sum_{m_x \in X} m_x \times \vec{r}, \quad (4.2)$$

where  $m_x$  is a point mass in set  $X$ ,  $M$  is total mass, and  $\vec{r}$  is the position vector of  $m_x$ . As the saliency maps were in grayscale, it was relatively easy to compute CoM. However, there was a question of how to calculate CoM for human-text and vanishing point detection. In[82], visual attention on a region is enveloped with a Gaussian distribution, as the most attention is paid to the center and that envelope degrades further from the center. We adopted this concept and fit a Gaussian within human-text detection boxes and around the vanishing points we had detected (Fig.4.6). Very thin boxes with high aspect ratios were eliminated. The upper-left corner coordinates, width and height of each box in the image, were taken and a Gaussian was fit inside the box for each of the remaining boxes (Eq.(4.3) and (4.4)). The spread of the Gaussian was thus a parameter to be tuned according to each box. The value  $\sigma$  was adjusted according to the diagonal length of the rectangle. The Gaussian that was fitted to  $i$ -th box  $b_i$  is

$$\sigma = \sqrt{\text{height}_{b_i}^2 + \text{width}_{b_i}^2}, \quad (4.3)$$

$$G = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{(x - x_{CoM})^2 + (y - y_{CoM})^2}{2\sigma^2} \right\}. \quad (4.4)$$

Another concept that can be borrowed from physics is the representation of forces as vectors. The coordinates of the CoM can be transformed into a force vector. The vectors computed from the CoMs of different image information channels can represent visual forces caused by different visual sources, thereby modeling the push and pull. The coordinates of the CoM are converted into a vector where the origin is the center of the image. Our approach was to combine CoM vectors through a vectorial sum. The values of vectors were standardized between 0 and 100.

### 4.3 Evaluation

After representing each visual weight component in a vectorial force structure, an evaluation of how well these forces predict the visual center of an image in relation to visual balance can be undertaken. This section describes the evaluation setup and results for our study.

**Setup.** In line with the competing visual forces mentioned in[2], we obtained the CoM vectors for saliency, human, text, and VP detection. However, the way these visual forces interact were unknown to us. As first criteria for performance, we inspected the absolute value of the difference between participant visual center values and CoM calculated for different features,  $|\mathbf{center}_{\text{participant}} - \mathbf{center}_{\text{feature}}|$ . A sorted version of these differences was evaluated to determine whether the feature in question caused a drop in the difference measure. It was observed that a simple vectorial sum failed to capture the actual resultant vector, as the contribution of each element was not known with certainty.

This problem can be addressed through a linear regression system, as explained below. The performance measure utilized for prediction is the mean square error in a cross-validation setting.

The contribution of each feature can be explored through a linear model that combines the forces. The problem can be considered a regression problem where the dependent variable is the location of the visual center and the predictors are the vectors for the visual forces at play. In order to show the effect of each element for prediction, we devised a hierarchical regression method that started with a base model that included only saliency information and incorporated each additional feature gradually:

$$\text{(Model 1) } \text{visual\_center} = \beta_0 + \beta_1 \cdot \text{CoM}_{\text{Saliency}} ,$$

$$\text{(Model 2) } \text{visual\_center} = \beta_0 + \beta_1 \cdot \text{CoM}_{\text{Saliency}} + \beta_2 \cdot \text{CoM}_{\text{Human}} ,$$

$$\text{(Model 3) } \text{visual\_center} = \beta_0 + \beta_1 \cdot \text{CoM}_{\text{Saliency}} + \beta_2 \cdot \text{CoM}_{\text{Human}} \\ + \beta_3 \cdot \text{CoM}_{\text{Text}} ,$$

$$\text{(Model 4) } \text{visual\_center} = \beta_0 + \beta_1 \cdot \text{CoM}_{\text{Saliency}} + \beta_2 \cdot \text{CoM}_{\text{Human}} \\ + \beta_3 \cdot \text{CoM}_{\text{Text}} + \beta_4 \cdot \text{CoM}_{\text{VP}} .$$

The hierarchical analysis showed that addition of human and VP detection information demonstrated statistically significant ( $p < 0.05$ ) improvements to the base model, while text detection information had no significant effect.

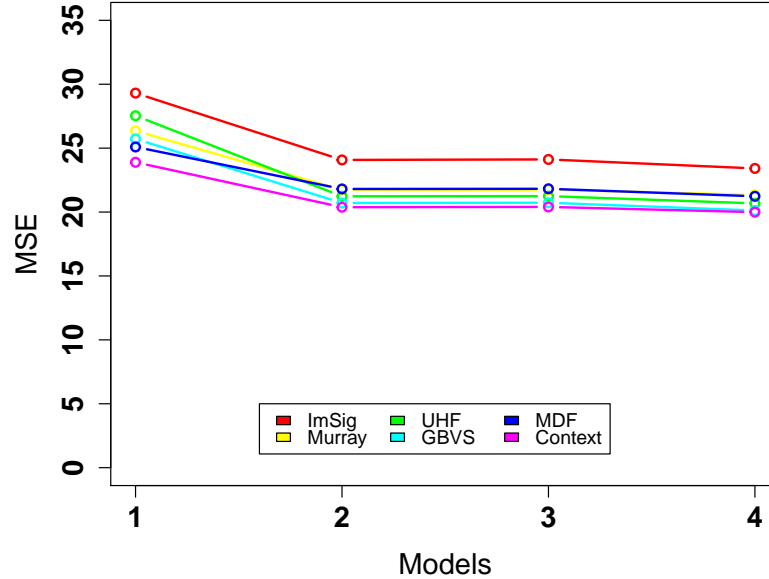


Figure 4.7: MSE values of Model 1, Model 2, Model 3, and Model 4 for selected saliency algorithms. Each line corresponds to a saliency algorithm chosen to represent the categories given in Table2.1.

The models were also tested with three-fold cross validation and the performances of the models were measured through mean square error (MSE) for each model. The results are shown in Fig.4.7.

**Results.** The analyses showed that saliency was useful but insufficient for the representation of compositional balance. Further analysis demonstrates that there is room for improvement. In light of empirical art studies, high-level features that might enhance saliency map performance were explored. The mean squared error (MSE) analysis showed that the addition of human information contributed to the representation power, as seen in the first two images in the first row of Fig.4.8. This contribution can be based on two observations. First, the humans that go undetected by saliency maps are sensed by including human detection information. Second, consolidating information regarding human positions helps suppress noisy saliency components.

The third, fourth, and fifth images in the first row of Fig.4.8 indicate why the dominant VP detection information helped boost prediction performance. Saliency

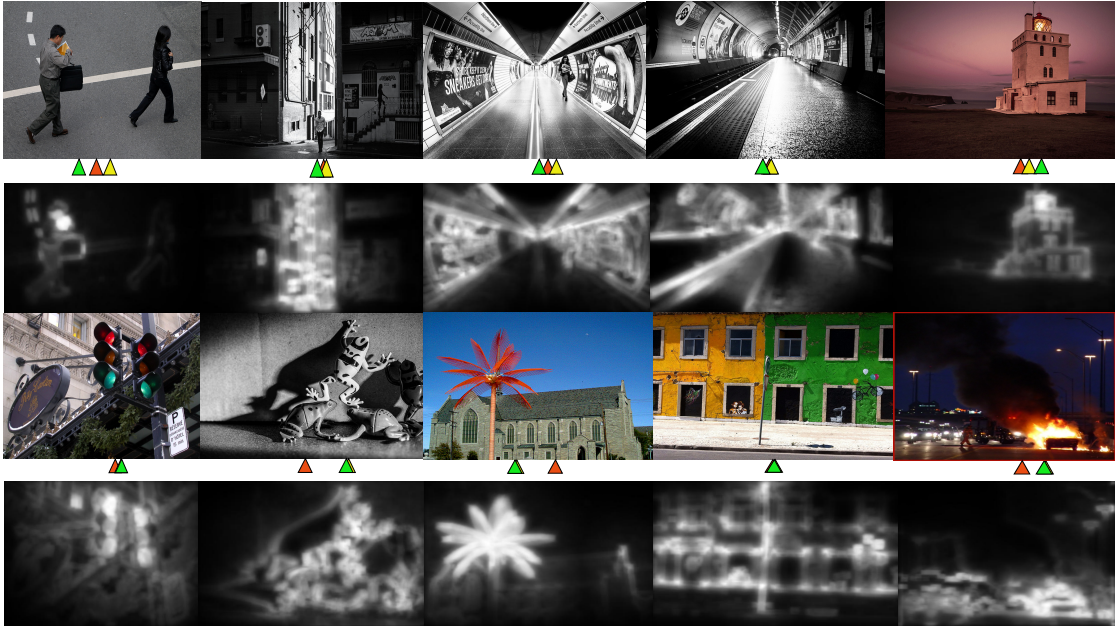


Figure 4.8: Evaluation on visual center prediction. The yellow triangle shows mean visual center position annotated by participants. The green triangle is the CoM of saliency images. The red triangle is the prediction of the proposed model. The top row demonstrates cases where our model improved prediction along with the saliency maps (context aware saliency). The lower row displays cases where the model was worse than saliency or didn't improve prediction.

maps fail to account for the effect of parallel long lines leading the eye toward one side of the picture, *i.e.* ceiling lines, the station floor, or the building edges in the case of the last image. The saliency values pick up on the components that make up parallel lines, which is mainly what bottom-up models are designed to achieve, but the values fail to capture the components' influence on balance. By including this information, the prediction performance can be improved.

On the other hand, including the text detection information did not provide any improvement in the base model. This finding can be interpreted in two ways. In the first case, the text information was not as important as we had expected, so saliency methods failing to detect text areas did not negatively affect the performance. The other case may be that the contrast difference around the text area and high-frequency texture of the text had already been captured via the bottom-up saliency method and adding the text location did not do anything more than stating the obvious.

Some cases in which our model did not improve prediction based solely on saliency or aggravated it are also shown in Fig.4.8 (lower row). Basically these are the cases where there were no dominant human or vanishing point components. The traffic lights in the first image in that row are the main object in the image that was captured by saliency. As there were no other components included in our model, the model predicted average response based solely on saliency information, which makes prediction performance worse. In the second image, the outline of the frog was captured by saliency, which is in line with participant ratings. The average response from our model makes it worse as the saliency CoM was multiplied with a coefficient. The third and fourth images have objects of different colors, *i.e.* a red tree and walls of different colors. The saliency map was able to capture the tree based on the color edges, while the model exacerbated the prediction due to regression coefficients. The contrast component, the fire, of the last image received more attention than the humans walking away from the car, which was not sensed by our model.

## 4.4 Discussion

Our study suggests that there are high-level visual elements that are influential in compositional balance but are not captured by bottom-up saliency maps. Accounting for these elements decreased the error compared to the use of visual center prediction employing only saliency maps. Inspection of the failure cases shows that there is room for further improvement. Compositional components that merit further consideration are explored during our study. These components are color and contrast. As seen in the lower row of Fig.4.8, the color of the tree and the color of the walls created different visual weights. Saliency captured the visual center correctly, as it sensed the edge structures of the tree and the windows. However, the color information was not encoded. It might be useful to account for the different visual weight of the different color areas [72]. The last image in the row suggests objects of contrast may have different visual weight.

In [79], saliency maps were learned from actual gaze maps, and it was concluded that humans, body parts, animals, and text receive more attention in an image.

As a top-down approach was adopted in the researchers' study, the contribution of each visual element was not studied. In our study, we attempted to analyze each element.

The hierarchical linear regression utilized in our study may have been effective in testing for the contributions of added high-level features. One aspect that may not be captured via simple linear regressions is the interaction between these features. Furthermore, linear regression may also fail to detect nonlinear relationship between the features and visual center locations given by participants. More complex learning schemes could be employed.

Another aspect of the study that could be improved is the size of the image dataset. The number of images is large compared to that used in empirical art studies, but if it were even larger, it would be more comprehensive. As there are no previous datasets related to this problem, the pictures had to be selected carefully to ensure that they showed balance characteristics. This process required the manual inspection of the pictures, which constrained the number of images substantially. Another point that should be made about the dataset collection is the online survey. As the study was conducted with a group of graduate students in a controlled lab environment, we deemed the detection of uninformative annotators unnecessary. If a much larger participant base were used, the elimination of uninformative annotators would need to be incorporated. In our study, the participants were from a small pool of graduate students whose demographics were not recorded. In the next stage of visual balance research, the influence of demographics could be further explored. In terms of different balance characteristics, the next step is an investigation of the association of the visual center with dynamic balance and imbalance.

In the future, the implementation of other contributing high-level features and the analysis of their effects on prediction performance would be helpful in pinpointing visual elements in the composition that could be used to strike a balanced structure. As visual center prediction improves, the seam carving, image thumb-nailing, and retargeting applications may perform better.

## 4.5 Summary

We investigated the relationship between compositional balance and the saliency concept through saliency maps' predictive power in relation to an image's visual center. In order to achieve this, we compiled a dataset of images that were rated by participants via an online survey. Different bottom-up saliency methods were run on these images. The centers of mass of saliency maps were directly utilized to predict the visual centers. In addition to saliency information, we included human, text, and vanishing point detection information based on theories in empirical art studies. The analyses demonstrated the addition of human and vanishing point detection to saliency improved the prediction of the locations of the visual centers of images.



# Differences and Biases in Affect Evoked by Visual Features in Latent Crowdsourcing Demographic Groups

## 5.1 Introduction

In this chapter, we consider the emotional aspect of visual features. The scope is set to the latent demographic groups on the popular crowdsourcing platform Amazon Mechanical Turk (AMT). The demographics of AMT workers have been investigated in previous studies. However, the workers' affective/emotional responses have not been reported. In this study, we analyze the agreements and differences in participants' emotional responses to visual features.

We start describing our method with the first section in light of context given in previous chapters. The data collection procedure is explained in detail. We follow with which visual features we extracted and why we are interested in them. We elucidate the process we cluster the images and participants of our study. In the ensuing section, the findings from the analysis of relations between visual features of images topics and latent demographic groups are provided. The chapter closes with a discussion of limitations of the work and possible directions.

## 5.2 Method

In order to capture the biases and differences in the affective responses of different crowdsourcing populations, we collect visual data including elicited emotions and aesthetic ratings from our crowdsourcing participant pool. AMT is utilized because it is one of the most popular crowdsourcing platforms. In this section, we expand on how the data were collected and analyzed.

### 5.2.1 Data Collection

**Image Dataset.** Collecting images from the Web is not a straightforward task. For our purposes of study, it was necessary to ensure that the images collected elicited affective responses on the part of the viewers. Metadata such as captions and comments provide information about what viewers thought or felt about a given particular image, which suggest that emotional queries can be used for image collection. The queries should be selected from words that define emotions. However, the emotional word space is grammatically complex and different words can be used to refer to the same emotion, causing redundancy. In an attempt to avoid redundancy, we adopted the exhaustive list of slightly more than 500 emotional words provided in [169] for our queries. The query set  $S$  may give an idea of the word set we used:

$$S = \{ \text{abashed, aghast, fatherly, frustrated, quarrelsome,} \\ \text{relieved, troubled, uproarious, violent, zealous, ...} \}$$

We collected 49,967 medium-sized ( $\approx 500 \times 500$ ) images from the photograph sharing site Flickr and ranked them according to an interestingness measure from the same site. Interestingness was incorporated in order to ensure the images elicited participants' emotional responses, as relying solely on relevance ranking returns neutral images. The interestingness measure helps to include emotional images without introducing any bias.

Although around 900 images per query were crawled, the number of images for each of the less common emotional words was lower. Images with only text (black and white) were omitted because they were considered non-natural. Images

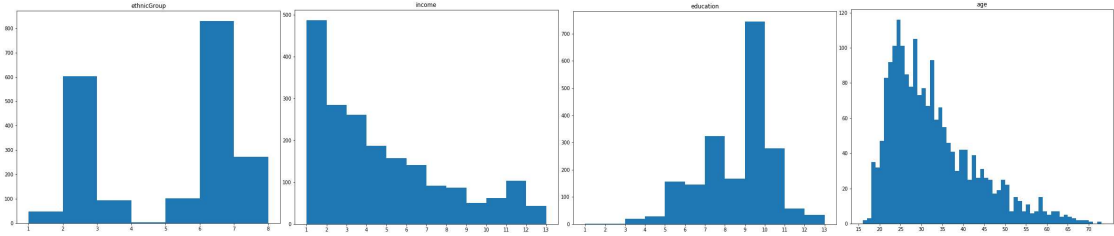


Figure 5.1: Ethnicity, income, education, and age distribution of AMT participants after data cleaning.

mainly consist of outdoor, indoor, landscape, and artistic themes that may or may not include humans.

**Emotional and Aesthetical Ratings.** Emotional scores and preference information were obtained through a user study on AMT. Design decisions for the survey included emotion representation, survey items, and AMT data quality measures.

There are two main types of representation for emotions: (i) categorical, and (ii) dimensional [170]. The emotions of users were captured via a set of categories and the users’ moods were measured prior to the experiment in [171]. Although the categorical approach is easy to understand, it is limited in cases of mixed emotions, as it is based on discrete emotion classes. In the dimensional approach, an emotion is represented with basic constructs such as valence, arousal, and dominance. This approach has been widely accepted by the computer science community as it is continuous and is convenient for use with classifiers [63, 172]. We thus chose to use the dimensional approach in this study. Aesthetic preferences were evaluated using the construct “likability,” which was measured along a Likert scale ranging from 1-7. The original International Affective Picture System (IAPS) study by [11] was rearranged for use within a digital survey environment. The Self-Assessment Manikin (SAM) was used to quantify emotions in [173]. We replaced the checkboxes and recorded dimensions using sliders. This allowed us to obtain relatively more continuous metrics and the manikins changed state as the sliders were moved. Participants in the study were first asked to view a consent page and then to fill out a demographic information page. Fig. 5.1 shows distribution of demographic attributes. The participants entered information about their gender, age, ethnicity, nationality, and income. The ethnicity groups of the participants were recorded using a pull-down menu with seven choices: (i) American Indian or Alaska Native,

(ii) Asian, (iii) African American, (iv) Native Hawaiian or Other Pacific Islander, (v) Hispanic or Latino, (vi) Not Hispanic or Latino, (vii) Other, which included an additional column in which the user was able to specify a value. Income information was obtained using a pull-down menu with choices that ranged from \$0 to \$100,000 in \$10,000 increments; two other options were also given: \$100,000-\$149,000 and \$150,000 or more. The level of education each user had attained was captured using a drop-down list. The categories were as follows: 1. No schooling completed, 2. Nursery school to 8th grade, 3. 9th , 10th, or 11th grade, 4. 12th grade, no diploma, 5. High school graduate – diploma or equivalent, 6. Some college credit, but less than 1 year, 7. 1 or more years of college, no degree, 8. Associate’s degree, 9. Bachelor’s degree, 10. Master’s degree, 11. Professional degree, 12. Doctorate degree.

After completing the demographic information page, the participants were presented a tutorial page explaining the emotional dimensions, the aim of the study, and the interface they would be using. They were asked to adjust the manikins using sliders to reflect how they “*actually felt while observing the image.*” They were told what each end of the sliders meant. The participants were shown the survey page seen in Fig. 5.2 after viewing an image for 6 seconds. On the survey page, the participants could choose to view the image again or they could go back to the tutorial page if necessary by using the relevant help buttons.

**Human Subject Study.** In this study, AMT was utilized to capture the affect and aesthetic ratings from Turkers. As previously mentioned, the main problem facing researchers who rely on AMT is the quality of participants’ responses. Workers’ potentially low pay has led to some concern over their dedication. Thus, their work results need to be controlled. AMT gives requesters the right to decline a worker for a particular task if he or she does not satisfy all of the requirements. The literature separates declined users into two categories, *speeders* and *cheaters* [30]. Speeders try to obtain the reward in a short amount of time using the least amount of cognitive effort. They might do this by giving the same response for each of the items (straight-lining). Cheaters attempt to optimize their effort by giving dishonest information to obtain the reward. There are different measures for spotting these users, including using measure of involvement (time spent) and attention check questions. It is claimed that Turkers can become so familiar with

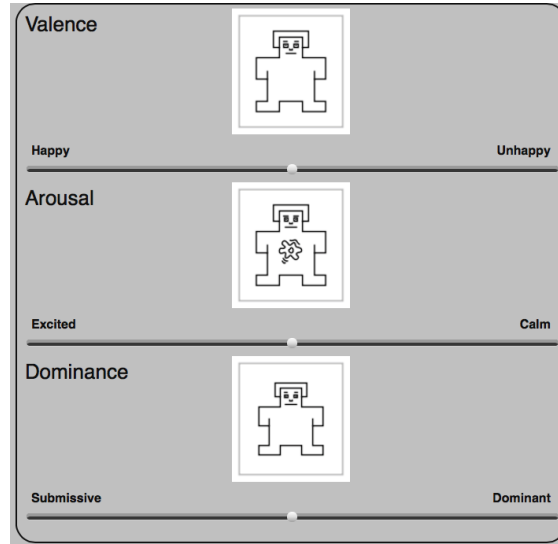


Figure 5.2: Survey Webpage used for human subject study [11]. SAMs are mapped to slider values so that users can employ the sliders to record the values that reflect their emotions.

attention check questions that they continue to provide irrelevant data even though they have passed the attention check [174]. As it is time-consuming to check every task completed, we devised a mechanism that measures users’ involvement. The survey we used automatically approved a user’s work if the user had spent a reasonable amount of time rating each image. The “reasonable amount of time” was determined during a pilot study that was conducted among a group of undergraduate students. The pilot study showed that rating 200 images takes approximately an hour. We record the total time a participant spends taking the study to determine the amount of effort the participant expended. In addition to measure of participant involvement, we provided a 200-image task in batches of 25 images to maintain workers’ attention. The workers were paid \$1.4/hour for the span of time it took them to finish the task.

**Post Cleaning.** As we only checked for the involvement of the participants during the survey, we applied a post-cleaning algorithm to eliminate spammers in the data set. For this task, we selected one of the state-of-the-art methods based on probabilistic graphical models, as described in [27]. Their method is a probabilistic multigraph approach to consensus modeling. The agreement between individual subjects is modeled (multigraph) using two latent variables, such as the subject’s

Category	Name	Description
<i>color</i>	Use of light	Mean brightness
	Pleasure, Arousal, Dominance	Valence, arousal, and dominance values calculated according to regression model.
	Colorfulness	EMD between the hue histogram of the image and uniform distribution.
	Color Names	Percentage of black, brown, blue, gray, green, orange, pink, purple, red, white, and yellow.
	Itten Contrasts	Contrast of brightness and saturation according to Itten's color space.
	Harmony	Harmony of colors according to Itten's color accordance.
<i>texture</i>	Wavelet Texture	Daubechies wavelet coefficients for HSV color space for levels 1-3 and their summation for each channel.
	Gray-level Co-occurrence Matrix	Features of contrast, correlation, energy, and homogeneity based on the co-occurrence matrix.
<i>composition</i>	Low Depth of Field	The ratio of the amount of detail in the rectangle encompassing the image center to the amount of detail in the center section of the image in the whole image for each channel.
	Rule of Thirds	Average hue, saturation, and value for the center rectangle.
<i>shape</i>	Roundness	How much the image is dominated by round structures.
	Angularity	How much the image is dominated by concave structures.
	Complexity	Number of segments in the image.
<i>dynamics</i>	Percentage of edge pixels	The percentage of edge pixels representing how dynamic the picture is.
<i>others</i>	Aspect ratio	The ratio of the width to the height of an image.

Table 5.1: The affective and aesthetic features utilized in gender difference analysis.

reliability (Bernoulli distribution) and the extent to which the particular subject's response agrees with other reliable responses (Beta distribution). The parameters for these latent distributions are obtained through an expectation maximization (EM) framework.

### 5.2.2 Visual Features

In order to relate images and emotional responses, images need to be characterized by relevant visual feature extraction. Features are chosen from those that have been commonly used in affective computing and computational aesthetics. While these features capture color, illuminance, and texture characteristics, they also provide information about the photographic and emotional properties of a picture [19, 67, 175, 176]. The extracted features are given in Table 5.1.

**Color percentage.** The make-up of an image in terms of discrete colors often used in daily life.

The colors employed in affective classification are obtained through a model-learned over a color name dataset [67]. Eleven different colors are returned, (Table 5.1).

**Colorfulness.** The Earth Mover’s Distance (EMD) between uniform color distribution and the image’s color distribution is computed.

**Harmony.** How much the image adheres to color accordance theory [67]. The dominating colors in the hue histogram are mapped to this wheel as points. The similarity between the inner angles of the polygon among these points and those of a regular polygon with the same number of vertices shows harmony adherence.

**Pleasure, arousal, and dominance.** Emotional values obtained through the regression relation given in [67].

**Spatial graininess-smoothness.** The normalized sum of the Daubechies wavelet transform coefficients at three levels is computed for HSV channels.

**Gray-level Co-occurrence Matrix Features.** This feature set is one of the most popular feature sets for describing texture. As suggested by [67], we considered the *contrast*, *correlation*, *energy*, and *homogeneity* of an image.

**Low Depth of Field.** A compositional effect that is commonly utilized to put focus on the main object, while making the background blurry and less attention-getting. The center area of an image has more detail than the other areas of an image. The ratio of wavelet coefficients in the center rectangle in a 3x3 grid of the image to the coefficients of the rest of the areas.

**Rule of Thirds.** A compositional practice in photography where the main object is placed within or at the border of the center rectangle in a 3x3 grid of the image. The average of hue, saturation, and value within the rectangle are computed to represent this rule.

**Dynamics.** The dynamic properties of images are characterized by computing the edge pixel percentage.

**Roundness.** The definition of “roundness” is “a measure of how closely the shape of an object approaches that of a circle” [175]. In the case of images, the measure of roundness captures how much the image contains round objects.

**Angularity.** The dictionary definition of “angularity” is “the quality of being angular” [175]. However, there are various approaches to the definition of “angular.”

Interviews with five human subjects yielded keywords such as concave, *sword-like* structures and tall, lean, and bone-shaped structures.

**Simplicity-Complexity.** Simplicity is analyzed and represented using two approaches: (i) Minimalistic structures that are used in a given representation; and (ii) the simple organization of these structures. In light of these, simplicity-complexity is defined as the number of segments with the same image segmentation approach [175].

**Other Features.** In addition to the aforementioned features, we investigate several other visual cues that could influence affective ratings. For example, [19, 171] argued that the size and aspect ratio of an image may influence its ratings.

### 5.2.3 Data Analysis

Ultimately, we collected ratings for a total of 41,255 images from 2,063 workers. Of these workers, 1,094 were female (coded as 1) and 969 were male (coded as 0). As seen Fig. 1, the two dominant ethnicity groups were “not Hispanic, not Latino” and “Asians.” Income levels were skewed towards participants who have income levels less than \$10,000. Most of the participants had bachelor’s degrees or had one or more years of college experience and no degree (i.e., undergraduates). The ages of the participants ranged from 16 to 72 with a skew towards the age of 25. The dominant age group was between 25 and 35.

‘*Valence*’ had a mean of 5.7 with a standard deviation of 1.9, ‘*Arousal*’ had a mean of 5.2 with a standard deviation of 2.0, and ‘*Dominance*’ had a mean of 4.9 with a standard deviation of 1.8. These values were computed for all emotion measures using a scale ranging from 1 to 9. In Fig. 5.2, ‘*happy*’ means a higher valence score; ‘*excited*’ means a higher arousal score, and ‘*dominant*’ means a higher dominance score.

#### Latent Demographics Group

Our analysis of the complete dataset separated according to one demographic variable did not provide any statistically significant results in terms of differences as the other variables were not controlled for. The world is not linear and looking for linear relationships through regression analysis may not be suitable for this



type of work. We need an analysis that simultaneously accounts for multiple demographic variables to characterize different demographic groups. This calls for a scheme that groups similar people in a single category, which is the basic definition of clustering. A demographic cluster can be defined in the same manner as we capture the topic of a document, which is by using a group of keywords such as “young low-income graduates.” This similarity can be identified by employing a topic modeling approach to obtain latent abstract groups in the Turker population sample that we obtained.

One of the topic modeling methods, Latent Dirichlet Allocation (LDA), is a generative probabilistic model of a corpus. In this approach, documents are represented as random mixtures over latent topics, which are portrayed by a distribution over word [177]. LDA was employed in [176] to characterize the favorite images of individuals with certain personality traits. Using a person’s favorite images, the researchers created abstract topics of images to obtain the average images for each abstract topic. These average images were used to train a model to predict the personality traits of individuals, given their favorite images. In a similar manner, our purpose is to characterize abstract groups out of samples. Hence, we applied LDA to our participant pool. With the exception of age, the demographic information for each user (i.e., ethnicity, education, income) were already quantized. If we imagine each person as a document, then the quantized demographic values are word frequencies. Applying LDA topic modeling to the study’s participants, we expected to find that similar people belonged to a single demographic theme. There were four such groups obtained through topic modeling (Fig. 5.3):

- **Group 0:** The first group consisted mostly of individuals older than 35. The individuals in these groups mostly identified as “not Hispanic, not Latino.” They mainly had master’s degrees and annual incomes generally less than \$90,000. This group included more female participants (250) than male participants (140).
- **Group 1:** Subjects’ ages mainly fell between 30-35 in this group. This group was also dominated by the “not Hispanic, not Latino” ethnicity category, with the second-most prominent group being Asians.

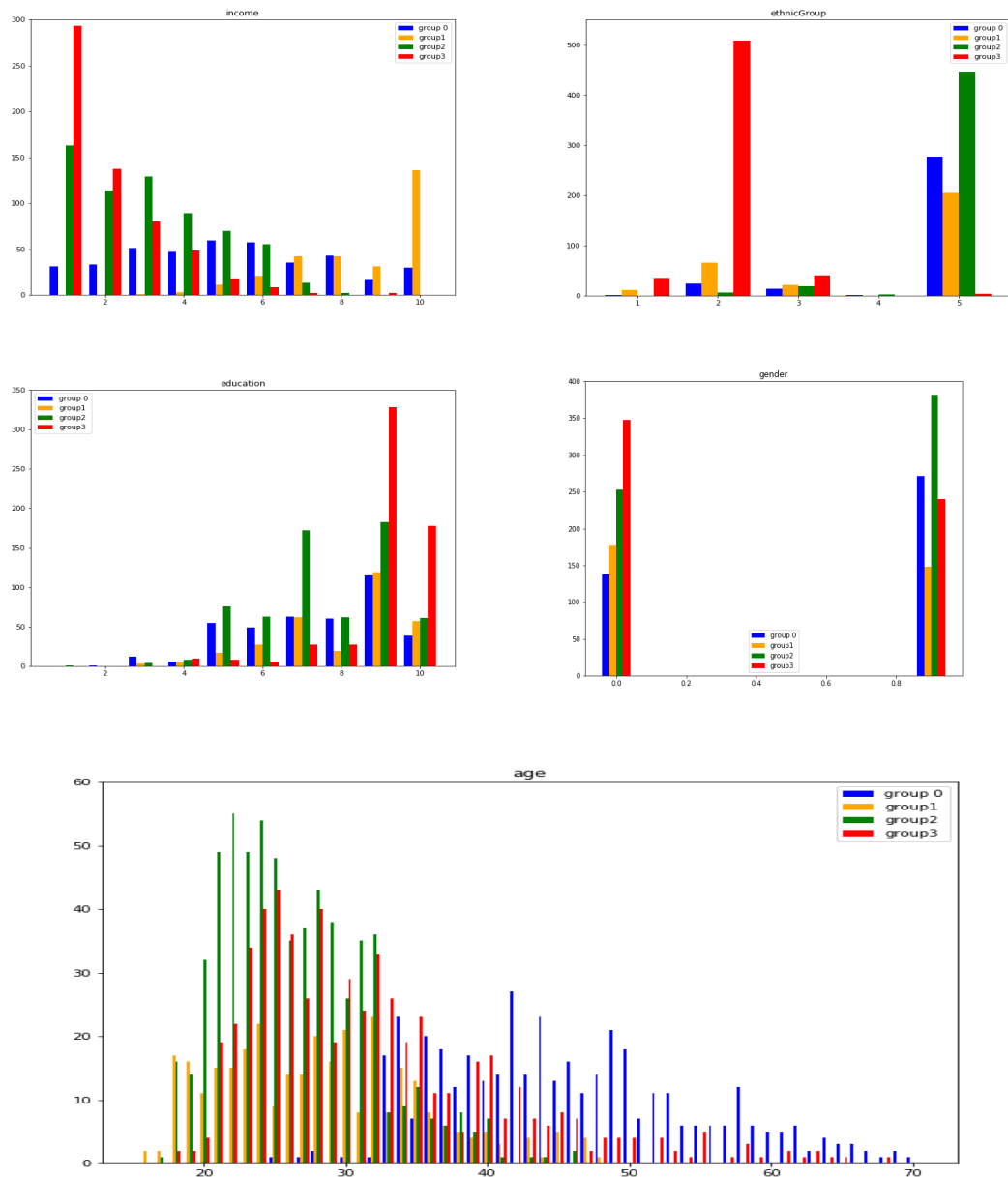


Figure 5.3: Latent representative demographic groups and their make-up in terms of income, ethnic group, education, gender, and age.

Individuals in the group mainly had bachelor's degrees with income levels ranging from \$100,000-\$149,999. This group had an equal number of female and male participants.

- **Group 2:** The third group consisted mostly of individuals between 19 to 32. The predominant ethnicity was again “not Hispanic, not Latino.” The individuals in the group largely had either bachelor’s degrees or one or more years of college, no degree. Incomes were generally less than \$70,000.
- **Group 3:** The fourth group also consisted of individuals ranging in age between 25 and 30. These individuals were mainly Asian males. These subjects generally held bachelor’s degrees and some held master’s degrees, too. They had incomes of less than \$40,000, which is in line with the Indian user profiles described in [31].

### Analysis of Responses to the Pictures

After determining the latent demographics groups in the sample, we analyzed the participants’ responses to the images to see whether there were correlations between visual features that may indicate affective biases and differences. As the results above indicate, selecting one feature and looking at the correlations between emotional constructs did not yield robust results. This may be because these features coexist in an image and together elicit a response. Hence, we had to further clump together the images that were rated by each group according to the images’ visual similarities in order to pin certain visual features to certain emotional responses. In a similar fashion to [176], we utilized LDA to separate the images into latent groups for each demographic group. The visual feature values were quantized. The quantized data were separated into topics that were characterized by feature names through LDA. For instance, a topic might be *blue, green, orange, gray, brown, red, yellow, percentage of edge pixels, purple, use of light*, where the order of the features’ names denotes the contribution of each particular feature to that topic. One example topic is demonstrated in Fig. 5.4.

It was observed that color features were more dominant in creating the visual topics because the dataset was composed of mostly color images. The same topic analysis was run without the color features to determine the role that other features play in representing visual topics. Correlation relationships between each emotional and aesthetic construct and each individual feature among the topics

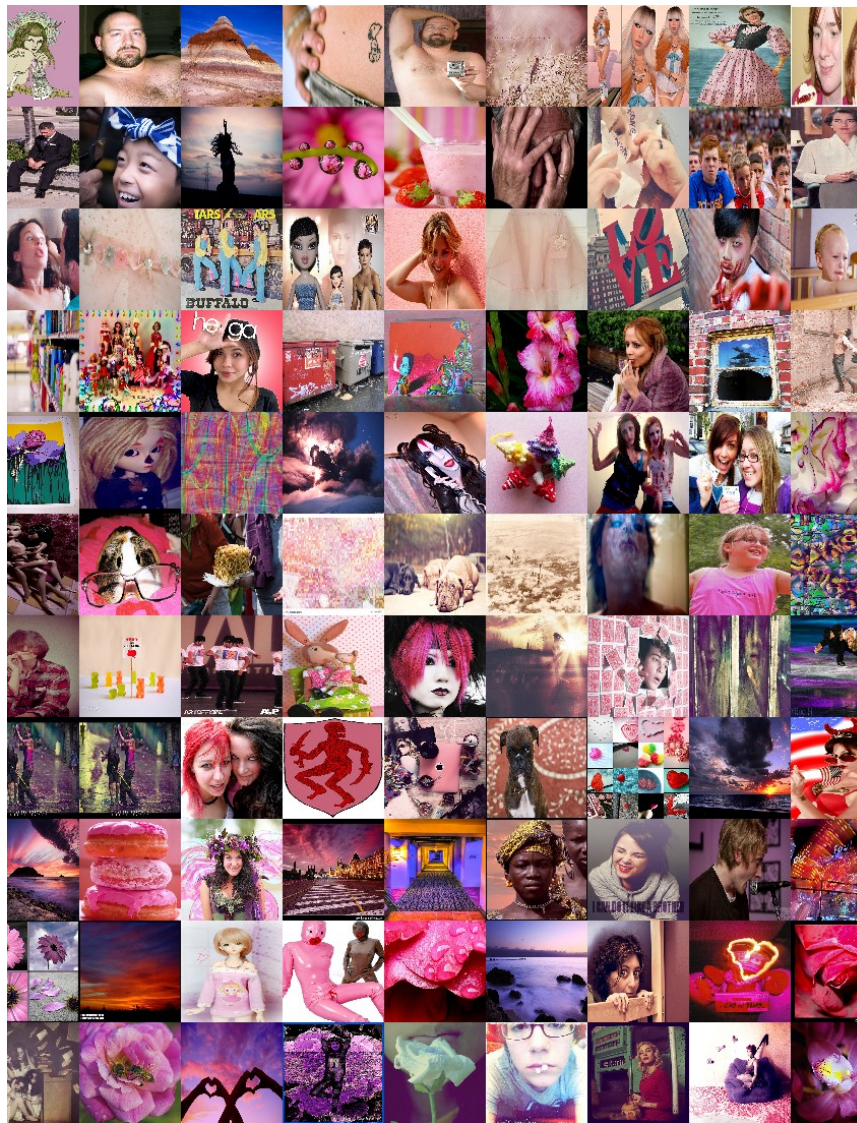


Figure 5.4: A sample of a topic whose most dominant words were purple, red, and white. Each image that belongs to this topic is ordered via the vectorial norm of the three features' values. Lower-triangle images demonstrate these features better.

were analyzed, and the Spearman correlation coefficients ( $\rho$ ) were calculated with associated statistical significance values ( $p$ ).

### 5.3 Findings

In this section, we provide the results of our correlation analysis for each demographic group. We discuss the commonalities and differences after describing our response analysis of each group.

**Group 0:** Group 1 had few significant correlated responses between some of the visual features. The use of light, which was average over the value channel of the HSV space, was negatively correlated with valence, but was evoked higher dominance for the emotions. The more overexposed the images were, the less valence the participants felt towards the images. As the complexity of the images increased, the images elicited more positive feelings among this group's participants. Images that had angular components evoked more positive feelings; however, this effect was weak. Brown and orange colors provoked positive feelings, whereas green, pink and white were associated with more negative emotions. The complexity of the images was positively associated with more energetic emotions. As the percentage of yellow, red, orange, and brown colors increased, more energetic emotions were elicited. Green and gray colors were related to calmer emotions. The white, orange, and red color percentages were positively related to the dominance of the emotions. Angularity and the gray and purple color percentages were associated with less dominance. Individuals belonging to this group aesthetically favored green and yellow colors. However, as the purple and orange color percentages or the complexity of the images increased, likability decreased.

**Group 1:** The complexity of the image and the blue color percentage were related to more positive emotions, whereas the brown, orange, and red color percentages and edge pixel measure were associated with more negative feelings. For individuals in this group, brown, gray, orange, white, and purple color percentages and complexity elicited more energetic emotions. The pictures that adhered to the rule of thirds were related to more energetic emotions. The color green and the percentage of edge pixels evoked calmer emotions. The color red, complexity, angularity, and adherence to the rule of thirds were associated with less dominance in the emotions evoked. Green and textural homogeneity in saturation were affiliated with more dominant emotions. Yellow, orange, and complexity were related positively to likability, whereas purple was negatively related.

**Group 2:** The green and white color percentages and the complexity of the images were positively linked to more positive emotions. However, brown, pink, orange, blue, yellow, and purple were associated with more negative emotions. The yellow and red color percentages and complexity were affiliated with more energetic emotions. Brown, gray, and colorfulness were correlated with calmer emotions. The color blue, edge pixel percentage, and textural homogeneity in saturation were associated with more dominant emotions, whereas complexity, the use of light, gray, purple, angularity, and the standard deviation of brightness were related to less dominant emotions. The purple, orange, pink, and blue color percentages were affiliated with less likability, whereas complexity and the gray color percentage were associated with more likability.

**Group 3:** The color percentages of brown, green, white, and red were related to more negative emotions. The percentages of yellow and purple, as well as angularity, were linked with more positive emotions. Brown, orange, red, and purple color percentages and complexity were associated with more energetic emotions, while it was the opposite case for green and white color percentages. White, orange, gray, and yellow color percentages, as well as textural contrast and homogeneity in saturation channel, were more related to dominant emotions. However, blue and purple color percentages, complexity, and level of detail were affiliated with less dominant emotions. For this group of individuals, the color orange, angularity, level of detail, and complexity were more positively linked with likability, whereas the color red and textural contrast for saturation channel were negatively linked.

### 5.3.1 Commonalities Across Groups

Groups 1,2, and 3 showed similarly negative correlations between the color brown percentage and valence. Groups 0, 1, and 2 demonstrated parallel valence responses to the complexity of the images. Groups 0 and 3 had similar valence-green, valence-white, and valence-angularity relations. Groups 0 and 2 both had a negative pink-valence relation. Groups 1 and 2 both demonstrated a negative relation between the color orange and valence. Both Groups 1 and 3 demonstrated a negative relation between red color percentage and valence.

Groups 0 and 2 both showed positive arousal-yellow correlations. For the same

groups, gray had a negative relation to arousal. Groups 0, 1, and 3 demonstrated similar trends in their brown-arousal, green-arousal, and orange-arousal relations. Groups 0 and 3 showed the same positive correlation between the color red and arousal. The relation between purple and arousal was similar in Groups 1 and 3. All the groups exhibited similarly positive correlations between complexity and arousal.

In terms of dominance, the responses from all of the groups show parallelism in the textural homogeneity of saturation-dominance and angularity-dominance relations. Groups 0, 2, and 3 all demonstrated a negative relationship between purple color percentage and dominance. Groups 0 and 3 had similarly positive white-dominance and orange-dominance correlations. Individuals from Groups 1 and 2 had similar dominance response associations in terms of blue percentage. The percentage of edge pixels had a common positive relation with dominance in Groups 2 and 3.

### 5.3.2 Differences Across Groups

Group 0 diverged from other groups by demonstrating a positive correlation between brown color and valence. Group 2 distinguished itself from the others by exhibiting a positive correlation for white-valence and green-valence associations. Groups 1 and 2 differed in terms of the blue-valence relationship. Groups 2 and 3 differed in terms of the purple-valence correlation.

Group 3 diverged from other groups with a negative brown-arousal link. Group 1 differed from Groups 0 and 2 by demonstrating a positive correlation between gray and arousal. Groups 1 and 3 differed in terms of the white-arousal link.

Groups 0 and 1 differed in their relations between red and dominance. Group 3 had a different relation between blue and dominance than Groups 1 and 2. Groups 0 and 2 differed in their use of the light-dominance correlation. The percentage of edge pixels had different correlation relations with dominance among Group 0 and Groups 2 and 3. Group 1 differed from Group 3 in terms of the green-dominance relation. While Groups 0 and 2 both had negative links between orange and likability, Groups 1 and 3 showed positive relations. Groups 1 and 2 differed on the correlation between gray and likability.

## 5.4 Discussion

Our findings reveal interesting instances of agreement and difference across the study’s groups. One aspect of our study that merits discussion is the way we separated the participants into demographic groups. Other grouping methods such as clustering also exist. In our experience, however, the simplest clustering method, k-means, creates many small clusters that make analysis harder and intractable. Treating each participant as a document yielded meaningful separation.

In our user study, we only checked user involvement time as a quality check and relied on a computational method to eliminate spammers. We could have made use of intermittent attention check questions during the survey. [174] described how Turkers become accustomed to attention check questions; hence, they provide irrelevant data even though they have passed the attention check. As the performance of our computational data cleaning approach was better than acceptable, we chose to forgo the incorporation of attention check questions into the survey.

One other factor that may raise a question is why we chose not to compensate for chance in our correlation analyses. As our study involved more than a thousand correlation analyses, chance may have affected our results. One correction for chance is Bonferroni’s correction. This correction is more or less conservative, and any correction may leave out significant results that may be interesting to the audience. Hence, we decided to provide the results as they are.

## 5.5 Summary

In this chapter, we investigated emotional agreements and differences in response to visual features across demographic groups of a crowdsourcing participant base. A diverse pool of participants was recruited from around the world through AMT. Latent demographic groups were extracted, and we inspected each group’s ratings of image sets. The images were passed through topic modeling to identify the clustered visual features that existed in our dataset. Correlation analyses between visual features and emotional-preferential constructs were conducted. We found that the groups had different responses to certain colors, whereas they were in agreement in their responses to structural features such as simplicity-complexity.



# **Microtubule Segmentation and Guard Cell Pair Isolation**

## **6.1 Introduction**

This chapter provides details about our proposed tubular structure segmentation method discussed in Chapter 1. The proposed scheme to integrate SOAX and tensor voting, which were explained in Chapter 3 is described in more detail in this chapter.

The chapter starts out with giving the details of our method. This includes a recap of Frangi filter and vessel enhancing diffusion for 2D. We continue with the introduction of improved curve initialization for MT detection. Then, we provide how we combine tensor voting information into stretching active contours framework (SOAX). The chapter goes on with a validation survey and generalizability experiment. Finally, we provide a discussion about the limitations of the work.

## **6.2 Method**

In this section, we provide the details on our framework. Firstly, the tensor voting, vesselness and SOAC are explained and how we change the energy functional is provided.

### 6.2.1 Frangi Vesselness Measure for 2D

[40] proposed vessel enhancement as a filtering process that searches for geometrical structures (tubular structure, MTs in our case). The second derivative of a Gaussian kernel ( $G$ ) at scale  $s$  is a way to examine the contrast between the regions inside and outside the range  $(s, -s)$  in the direction of the derivative. An analysis of the eigenvalues of Hessian given in Eq.(6.1) provides geometrical information regarding the image.

$$\mathcal{H} = \begin{bmatrix} I_{xx} & I_{xy} \\ I_{yx} & I_{yy} \end{bmatrix}, \quad \text{where } I_{xx} = \frac{\partial^2}{\partial x^2} G(s) * I. \quad (6.1)$$

The Hessian matrix,  $\mathcal{H}_{o,s}$ , of the image is calculated in  $x_o$  with scale  $s$ . Using the eigenvalue analysis of Hessian, we are able to obtain the principal directions in which the local second-order structure of the image can be decomposed in 2D. For the ordering of eigenvalues as  $|\lambda_1| \leq |\lambda_2|$ , a pixel belonging to a vessel region has  $|\lambda_1| \approx 0$  since the second-order characteristics have a small change in the direction of the vessel, and  $|\lambda_1| \ll |\lambda_2|$ , when  $\lambda_2$  is negative, as the vessels are light tubular structures against a dark background in CT scans. Using the values of  $\lambda$ 's, we can determine the kind of region in which the pixel is situated *i.e.*, blob, tube, or none in 2D. Two ratios are defined as in [40],

$$R_B = \frac{|\lambda_1|}{|\lambda_2|}, \quad S = \sqrt{\sum_{i \leq D} \lambda_j^2}, \quad (6.2)$$

where  $D$  is the dimensions of the image. The ratio  $R_B$  compares the eigenvalues in the directions of the least and the most change. The more the ratio tends toward 0, the more likely the pixel/voxel is part of a line-like/tubular structure. For blob-like structures, the eigenvalues are expected to be close to each other because the intensity change in each direction is similar in magnitude. As a result, the ratio approaches 1 for blob-like structures (Table 6.1). The parameter  $S$  is the Euclidean norm of the eigenvalues and is called the second-order structureness. This parameter works to clean the noise in the image, accounting for the strength of the second-order characteristics of the image.

A vessel, by having a circular cross section, has a plate-like characteristic per-

$\lambda_1$	$\lambda_2$	Structure
Noisy	Noisy	noise
Low	H (-)	blob-like (bright)
High (-)	High (-)	tubular (bright)

Table 6.1: Structure analysis based on the eigenvalues of the Hessian matrix, when the eigenvalues are ordered as  $|\lambda_1| \leq |\lambda_2|$ . This classification applies to bright tubular objects against a dark background.

pendicular to the vessel axis. Using these two model characteristics, [40] suggested a vesselness measure for 2D using the parameters described above, which is

$$V_o(s) = \begin{cases} 0, & \text{if } \lambda_2 > 0 \\ e^{-\frac{R_B^2}{2\beta^2}} \left(1 - e^{-\frac{s^2}{2c^2}}\right), & \text{otherwise} \end{cases} \quad (6.3)$$

where  $\alpha$ ,  $\beta$ , and  $c$  are constants used to control the sensitivity of the parameters. The vesselness measure approaches 1 as the structure becomes more tubular.

### 6.2.2 Preprocessing

Confocal microscopy images have a low signal-to-noise ratio (SNR), as mentioned in Chapter 3. We adopt two different denoising algorithms in order to eliminate noise in our data before obtaining 2D maximum intensity projections. Low-rank denoising was considered to be a suitable candidate for our initial denoising step. With this method, reverting to a low-rank subspace might eliminate most of the noise, but it would also sacrifice some of the original signal in the data mixed with error (*e.g.* noise).

We utilized the OPTShrink approach given in Chapter 3 for initial denoising. Another denoising method that we utilized was a specifically designed diffusion tensor employed to enhance tubular structures in noisy images. This method was proposed by [141]. Denoising was achieved via a diffusion tensor that used the eigenvalue analysis introduced above. The diffusion tensor is a diagonal matrix whose diagonals are

$$\lambda_1 = 1 + (\omega - 1) \cdot V^{\frac{1}{s}}, \quad (6.4)$$

$$\lambda_2 = 1 + (\epsilon - 1) \cdot V^{\frac{1}{s}}, \quad (6.5)$$

where  $\lambda_1$  is the smallest eigenvalue of the Hessian matrix. In the equation,  $V \in [0, 1]$  and  $\omega > \epsilon$ ,  $\epsilon > 0$ , while  $s$  is a positive real constant. The parameter  $\omega$  denotes the strength of anisotropic diffusion.  $\epsilon$  should be a small value to represent the isotropic diffusion component.  $s$  controls the weight of the vesselness term. With this formulation of a diffusion tensor, diffusion is performed along the tubular structure in the same direction as that of the eigenvector associated with the smallest eigenvalue of Hessian. This way tubule walls are protected from diffusion, but noisy parts or non-structure parts are diffused and smoothed out.

### 6.2.3 Heuristics Initialization

The ridge detection scheme proposed by [46] works well when there is a good amount of contrast difference between the foreground and the background. The ridge detection introduces false positives (FP) because it causes the evolution to start with FP snakes. During the evolution, those FP snakes that are far away from other snakes shrink to non-existence, but those that are close to other snakes become joined through overlap check, causing jagged end snakes. In order to improve the initial snakes, we benefit from the structural regularity of MTs in stomata cell pairs.

The MTs are located along the radial direction from the center of a stoma. The intensity profile that is obtained by tracing the cell pair in the tangential direction demonstrates that the relative maxima coincide with tubular structures. We extract the relative maxima along the tangential direction and link them through edge-linking provided in [134]. The improvement is demonstrated in Fig. 6.1. Relative maxima detection in tangential direction can be formulated as

$$\max (I_N(\rho, \theta)), \theta = [0, 2\pi], \quad (6.6)$$

where  $I_N$  is the image neighborhood at given  $\rho$  and  $\theta$ , which are the polar coordinate values. The whole  $\theta$  range is iterated for a given  $\rho$ , which covers the distance from the inner wall to the outer wall.



Figure 6.1: Improvements benefiting from a heuristic initialization. The second column shows seed points through tangential local maxima as opposed to ridge detection in the first row.

#### 6.2.4 Tubular Geometry Aware SOACs (G-SOAX)

We propose integrating the benefits of tensor voting into the SOAX framework so that the curves extend through the gaps caused by imaging imperfections. The incorporation of tensor voting can be achieved through the manipulation of an external force component on the curves. Furthermore, we aim to follow the continuity of tubular structure, which stands out from the other curves in 2D projections of confocal micrographs, by enforcing vesselness the measure.

The Frangi framework’s measure of vesselness/tubularity can be incorporated into the tensor voting framework. The tensor initialization and voting framework can be adapted to accommodate tubular structures. In this study, we exchange the structure tensor encoder with the Hessian tensor, as we are interested in detecting tubular structures. The Hessians were computed at different scales, and the responses were summed over different scales to suppress noisy components. This is because tubule regions produce consistent responses while noise regions produce

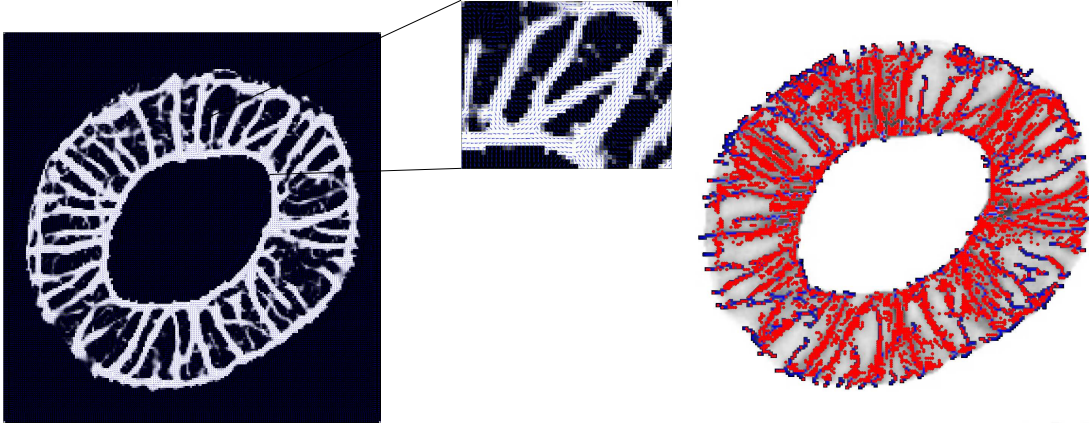


Figure 6.2: The principle eigenvector field of voting tensor  $\mathcal{T}$  (blue arrows) overlaid on top of the vesselness map, and curl points below the mean curl magnitude overlaid on the original sample in red. Blue dots represent initial curve points extracted through our tangential tracing heuristic.

random responses. After the accumulation, the tubule regions have a higher response rate than noise regions. Hence, the term  $\lambda_1 - \lambda_2$  from Eq.(3.44) shows tubular structure saliency. The principal eigenvector field of the voting tensor can be considered as the elongation direction of the detected, but broken tubular structures. The vector field flow is free of any turbulence and consistent with the tubule direction. This means that the curves that move along the areas where there is less turbulence in the voting vector field are desirable. The low turbulence areas can be quantified by the curl of the vector field. The curve should move in the direction of the low curl. Furthermore, we include a vesselness map to create a tighter fit, guiding the curves. The curve follows the high vesselness regions. Fig. 6.2 visualizes the field on top of the vesselness map and shows the low curl areas. Taking the new terms into account, the external energy factor to minimize becomes

$$\mathcal{E}_{ext}(\mathbf{r}) = \int_0^L k_{img} E_{img}(\mathbf{r}(s)) + k_{str} E_{str}(\mathbf{r}(s)) + k_{curl} |\nabla \times \mathcal{T}_1| - k_v V_s ds, \quad (6.7)$$

where  $\nabla \times \mathcal{T}_1$  is the curl of the principle eigenvector field of vote tensor  $\mathcal{T}$  and  $V_s$  is the vesselness map value.

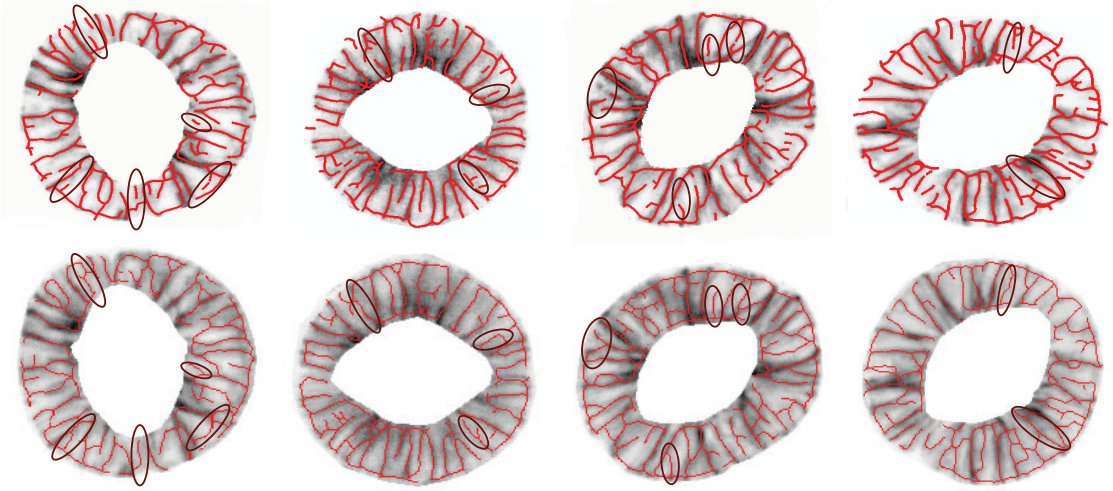


Figure 6.3: Comparison of normal SOAX (first row) versus curl and vesselness-driven G-SOAX (second row) in the same samples. The ellipses point to regions where G-SOAX continued evolving to combine the snakes.

### 6.2.5 Artifact Removal

Fig. 6.3 shows the results of SOAX in comparison to voting field curl and vesselness guided SOAX (G-SOAX). It is noted that there exist inner and outer wall artifacts due to faint tubule-like artifacts near the cell wall and between the tubules. In order to remove these artifacts, we devise an automatic morphology-based *dewalling* approach. In this approach, the resulting snakes are transformed into a polar space where tubules that are more or less aligned in the radial direction are vertical and the artifacts are horizontal. This is visualized in an image where rows are  $\rho$  and columns are  $\theta$  in the polar space (Fig. 6.4). The problem can be reduced to distinguishing between the vertical and horizontal structures in the transformed image. The orientation of structures can be inferred through the response to the vertical and horizontal difference image filters. Vertical structures give a higher response to the horizontal difference filter, whereas horizontal structures give a higher response to the vertical difference filter.

With this simplification, an edge detection filter, Sobel in X and Y directions, can be utilized. The X-filter response is higher (brighter) for vertical structures, whereas the Y-filter response is higher for horizontal structures. Observing these higher response regions reveals that horizontal parts are contained within bounding



Figure 6.4: The artifact removal problem becomes easier as the resulting snakes are converted to polar coordinates. The boundary artifacts and tubules separate into vertical and horizontal structures. The use of a Sobel edge filter applied to the binary image can delineate these structures. The bounding box of connected components provides an idea of the structure orientation. The magenta bounding boxes show horizontal structures in both of the thresholded horizontal and vertical difference images.

boxes with a high aspect ratio. The elimination of connected components with high aspect ratio bounding boxes produces the tubular structures that lie in the radial direction in the actual image plane.

## 6.3 Evaluation

### 6.3.1 The Microtubule Image Dataset

MT images were collected using a Zeiss Axio Observer microscope attached to a Yokogawa CSU-X1 spinning disk head with a 100X objective (1.4 NA, oil immersion). The MTs were visualized by tagging TUA5, a component of MTs, with mCherry, a red fluorescent protein. A 561 nm excitation laser and a 593/40 nm emission filter were used. Z-stack images were collected with a step size of  $0.2 \mu\text{m}$  in Z. To enhance the signal-to-noise ratio, images were first background subtracted and contrast enhanced using ImageJ. The Sliding Paraboloid algorithm with a rolling ball radius of 30 pixels was used for background subtraction. Saturated pixels were set to 0.4 percent for contrast enhancement.



### 6.3.2 Survey-based Evaluation

For this challenging dataset, obtaining good ground truth data is questionable due to the low contrast difference between the background and tubules and their intertwined nature. In order to rigorously evaluate the proposed method against the state-of-the-art SOAX method, we devise a validation scheme where the baseline and our results are compared by plant biology faculty and graduate students who are knowledgeable about MTs. In this scheme, two parts of our method are evaluated. The first part of our survey, in which we add the voting field and vesselness to improve continuity and regularity, is evaluated to determine its efficacy. The second part of the survey compares the overall quality of SOAX and G-SOAX after the artifacts are removed.

We developed an online survey that starts with an opening page, which gives general information about the study and investigators. The next page shows the instructions document, which informs each participant about the purpose of the study and explains the two parts of the survey and participant expectations for these parts. The user has a chance to come back to these instructions at any time during the survey. Once the user finishes reading the instructions, the survey starts. The webpage randomly selects a 2D maximum projection of a cell sample and shows overlaid 2D SOAX and G-SOAX results side-by-side without informing the user which results were produced by which method. Prior to the survey, the heuristic seed point improvement is applied to both SOAX and G-SOAX to improve regularity, *i.e.*, less branching of snakes; otherwise, it would be apparent to the user which method has been used for which results. The participant is explicitly asked to *click on the radio button under the image with more continuity and tubule shape adherence*. The second part of the survey considers the overall quality of both algorithms after artifact removal. At the end of the first part, the participant is informed that the second part will start. The same cell samples with removed artifacts are shown to the participant side-by-side after reshuffling, and the participant is asked to compare the quality of the results of overall tubule tracking and select the better one. The user is asked to *click the radio button below the image you think, the tubules were tracked more continuously and fully*. The survey was taken by one faculty member and eight graduate students from the

Method	Part 1			Part 2			Total		
	Open	Closed	Overall	Open	Closed	Overall	Open	Closed	Overall
SOAX	0.83	1.0	0.90	0.0	0.08	0.03	0.42	0.54	0.47
G-SOAX	0.17	0.0	0.1	<b>1.0</b>	<b>0.92</b>	<b>0.97</b>	<b>0.58</b>	0.46	<b>0.53</b>

Table 6.2: Detailed analysis of the survey in terms of the ratio of the number of samples. Numbers in bold indicate ratios where G-SOAX was preferred over SOAX. SOAX was chosen more in the first phase of the survey. However, our method was chosen with a dominating ratio for the survey’s second part, which considers the overall quality of the segmentation.

plant biology department at Penn State. We had fifteen closed and sixteen open samples, for a total of 62 samples shown to participants during both parts of the survey. It took 30-40 minutes on average for a participant to complete the survey.

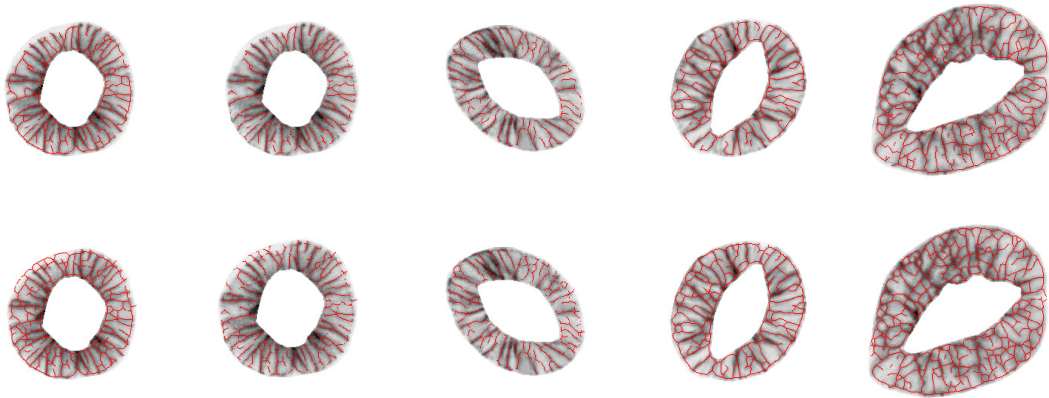


Figure 6.5: Some results from the online evaluation survey. The samples with boundary artifacts are from the first part of the survey and the cleared results are from the second part. The first row shows the G-SOAX results, while the second row shows the SOAX results. The first three columns are samples where the G-SOAX results were preferred, whereas the last two columns are samples where the SOAX results were preferred. It can be observed in the first three columns that G-SOAX helped to create more regular results free of unnecessary branches.

### 6.3.3 Results

In order to analyze the results of the survey, we adopted a consensus-based method. Since we framed the survey in a binary fashion (*i.e.*, one is or is not better than

the other), treating the preferences as votes and utilizing a simple majority vote count to decide which method was favored is appropriate. This is formulated in Eq.(6.8). That is,

$$M(s) = \begin{cases} 0, & \text{if } (\sum_{i=1}^n p_i) < \lfloor \frac{n}{2} \rfloor \\ 1, & \text{otherwise} \end{cases} \quad (6.8)$$

where  $M$  is the final preference vote result for sample  $s$ ,  $n$  is the number of people casting a vote for the sample, and  $p_i$  is the binary vote value from each participant, which is 0 for SOAX and 1 for G-SOAX. After the final votes are calculated, the preference percentages for the methods are computed. The results are summarized in Table 6.2. In retrospect, the results from the first part create perplexity in terms of continuity. A short debriefing with a few participants revealed that the pictures that were shown during the first phase with artifacts created confusion. However, the second part, which considered overall tubule tracing quality, revealed that our method was undoubtedly preferred. This points to confusion among the participants regarding the first part of the survey. Fig. 6.5 shows a few results where G-SOAX or SOAX were preferred. The G-SOAX results in the first column image have fewer false positive tracing tubules. This results in fewer crosses across the tubules, yielding less unnecessary branching. In the second column, the tubule tracing is more continuous in the G-SOAX results. However, the SOAX results are more broken and have false positive tracings at the boundary. The third column displays the results for a closed sample. Similarly, the G-SOAX result is more continuous and has fewer false positive tubule tracings. The SOAX results for the last two samples captured more tubules correctly than G-SOAX did.

### 6.3.4 Generalizability to Other Tubular Structure Data

In order to test the generalizability of our method, we consider other available 2D tubular structure datasets. The Cell Image Library supported by The American Society for Cell Biology has a good collection of biology datasets categorized by cell processes, cell components, cell types, and organisms. A dataset search using the query “microtubules” returns different types of tubular structure datasets. The availability of various tubular structure datasets can be considered another indication of the importance of the many different roles they play in organisms.

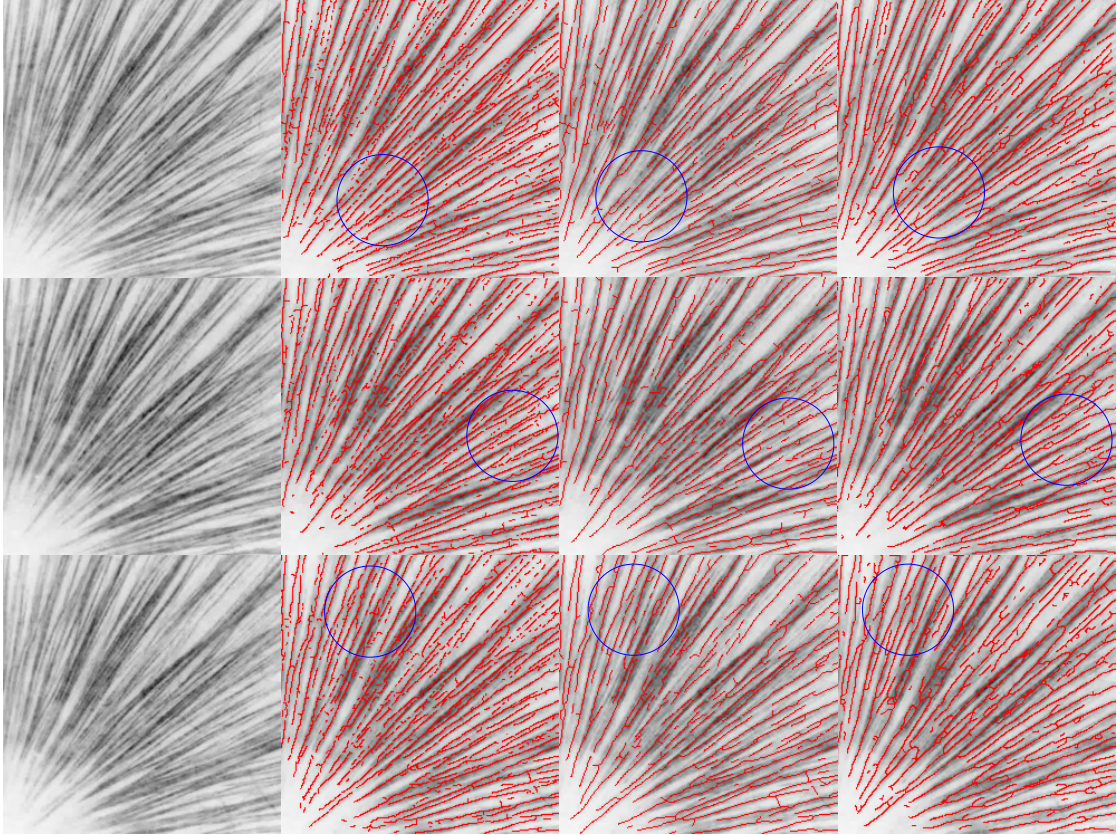


Figure 6.6: Comparison of the Aster data, where black structures are MTs. The first column shows the original images. The second and third columns are tubule tracing results for SOAX run at 100 and 10,000 maximum iterations, respectively. The last column shows the G-SOAX results. The blue circles point to exemplary regions where tubule continuity is better with our algorithm.

While looking for datasets with which to test our method, we establish a few criteria. First, the tubular structure organization should have some structural regularity, which helps with the initial seeding for the evolution process. Secondly, datasets that have tubular structures with similar intricacy and resolution, *i.e.* close MTs that are hard to distinguish, are preferred to set up similar challenges for our method. Datasets that already have grayscale images are favored as our method can be directly applied. One such dataset is provided in [178], where the growth dynamics of a radial array of MTs that are called *asters* was studied. Asters provide organization for larger cells in early vertebrate embryos. The study explored how asters at first grew in small embryo cells and then came to span the

whole cytoplasm, leading to cleavage. The MTs were imaged over time via total internal reflection fluorescence (TIRF) microscopy. Similar to MTs in stomatal guard cell pairs, asters are radially organized, emanating from centrosomes. The challenging part of the data is the thin tubular structures that have low contrast and are close to one another.

In the dataset, the centrosome is in the lower left-hand corner of the image, which provides an opportunity to utilize our radial initial curve improvement. Both SOAX and G-SOAX algorithms are run with a maximum of 100 iterations. For G-SOAX, the coefficients of the curl term and of the vesselness terms are both 1. The SOAX method is utilized with its own initialization method, which involves intensity ridges. It is understood from observations that a 100 maximum iterations limit is not sufficient to fully evolve SOAX; thus, 10,000 iterations are also run for that algorithm. For G-SOAX, it is observed that boundary artifact removal is unnecessary for this data. Fig. 6.6 shows the sample results for the first, the middle, and the last frames of a 320-frame aster growth video. As the tubules become thinner, it becomes harder to detect them. The continuity of some of the detected tubules is affected by low contrast for SOAX.

## 6.4 Discussion

The survey study results suggest that our method improves on the former SOAX framework in terms of overall tracing quality for biologists. An inspection of samples employed for the survey reveals that the tracing performance for closed cases are affected. This might be due to MT disintegration at the closed state of the guard cell (GC) pair [130, 131]. The radial organization of MTs breaks down at the closed state, and there is further MT disintegration. As seen in the last column of Fig. 6.5, the MTs have even less contrast compared to open cases. This presents a challenge for computer vision methods. Our method bundles a few tubules and considers them as a single MT due to the loss of contrast and inherent imaging artifacts, which produces false positives. In the future, a better denoising scheme that can resolve noise problems with both open and closed cases can improve the performance.

While our method shares similar parameters with SOAX, we eliminate the dependence on parameter  $\tau$  as the initialization scheme is changed. In stomata guard cell MT data, the center of the cell is automatically computed as the CoM of the cell pair. We keep other shared parameters such as viscosity ( $\gamma$ ) and radii for local foreground and background intensity computation at snake tips within the recommended intervals, as provided in [46]. We introduce two more parameters on top of the original parameter set. These are the coefficients of the tensor voting curl ( $k_{curl}$ ) and vesselness ( $k_V$ ) terms in the energy equation. These parameters influence the sensitivity of the evolution to these terms. As the curl term coefficient increases, the snakes follow the low curl areas in a stricter fashion. A higher  $k_{curl}$  makes the snakes follow very low voting curl areas, which might force them to stop before merging. As the vesselness term aims for higher values in the vesselness map, lower  $k_V$  enforces tighter evolution for higher vesselness, which makes the snakes more vulnerable to breakage. One general problem with evolutionary methods is that the initialization of seed curves or surfaces may affect the performance. It is the case in our work that the radial initial curve search improved the performance of SOAX. However, this might not be enough for another dataset with entirely different geometric patterns, which might interfere with our method’s performance. In a similar fashion, a severe imaging artifact may disrupt the evolution of snakes as it has a direct influence on external evolution energy.

Our artifact removal method may have worked for the MT positions. However, quantifying individual MTs may require a more structured approach. The current approach may remove artifacts that lie in the tangential direction; hence, cases in which actual tubules lie in the tangential direction can be problematic for this method. One possible research direction might be the alteration of the energy optimization step or the overlap check step. The ideal case would incorporate a specific penalizing term that avoids wall structures that cross too many tubules. Another approach might be to analyze the intensity profiles of those snake parts that run from one junction to another junction. The parts that show rather erratic intensity profiles can be removed.

Future work includes extending our approach to 3D volumes. The formulation of tensor voting in higher dimensions is available. One improvement might be the integration and implementation of 3D tensor voting with the 3D volume SOAX

framework. Another option is to keep the core algorithm for 2D as it is but to save the slice indices for each pixel in the memory while obtaining the maximum intensity projection images. Upon completion of tubular structure detection, the pixels detected can be put back into a 3D volume according to each pixel's slice index. The framework proposed above can be added to the end of the learning pipeline to refine the segmentation. As more labeled data are available, learning-based methods can be better leveraged for segmentation and tracing problems. For example, in the future, we can utilize deep learning to have more abundant visual representations from big data created by experts [179, 180]. These features can be directly used in an end-to-end deep learning framework or can be fed into another machine learning algorithm such as the Support-Vector Machine (SVM) to obtain the segmentation results.

## 6.5 Summary

We proposed a new tubular structure detection framework, G-SOAX, for biological images. Our method extends and improves upon a state-of-the-art active contour-based method, SOAX. With the proposed method, a structure/geometry sensing tensor voting scheme is integrated into the energy functional through curl information. Furthermore, the well-known Frangi vesselness measure is integrated to lead curves along the tubules. We tested our method on 2D maximum intensity projection images obtained in the guard cells of *Arabidopsis thaliana* seedlings via spinning disk confocal microscopy of fluorescent protein (FP)-labeled MTs. It was shown that the method is able to connect the tubules that seem to be broken due to contrast issues. This demonstrates the positive effect of visual continuity enforced by tensor voting and strengthened by a vesselness map in tubular structure detection for overall continuity.

# **3D Inner and Outer Wall Extraction of Stomata Guard Cell Pair Through Active Surfaces**

## **7.1 Introduction**

As mentioned in Chapter 1, being able to model a physical entity gives the scientist the power to predict and manipulate that entity's behavior. In plant stoma research, this leads to a faster iteration of the scientific process, while unveiling the dynamics of a stomatal complex. An aspect of modeling stomata is obtaining 3D surfaces during opening and closure interval and match the physical parameters of the model to these 3D surfaces. Hence, the geometry of stomata guard cells can be predicted for different pressures, temperatures and light conditions. In this chapter, we describe a software package and a procedure that can be utilized to obtain 3D surfaces of the inner or outer wall of stoma guard cell pairs.

The chapter starts out with a detailed description of our approach. Each step provides the problem they address and how we solved those problems. We provide a comparison of 3D models extracted via our method to other stoma guard cell pair models. The chapter ends with a brief discussion of the limitations of the proposed work.



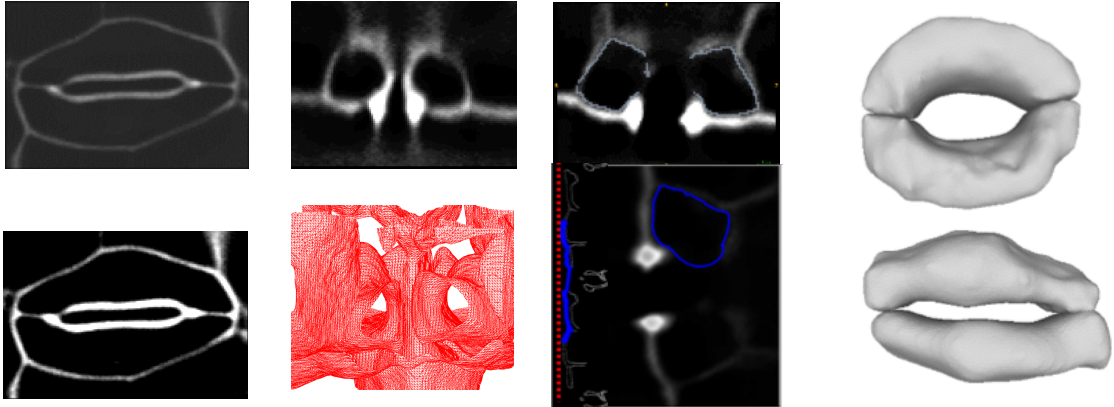


Figure 7.1: The pipeline of our method. The first step is manual contrast enhancement, which is followed by initial active surface segmentation. The third step is the extraction of the inner and outer walls while closing the gaps due to imaging methods. The final step involves the manual manipulation and rendering of the extracted surface into meshes for visualization.

## 7.2 Method

Fig. 7.1 shows our process of surface segmentation along with its challenges. Our method is composed of five parts that are preprocessing, initial surface extraction, inner or outer wall separation, averaging smoothing, and manual manipulation.

### 7.2.1 Manual Contrast Enhancement

Penetration of fluorescent material utilized for microscopy imaging worsens as we go deeper into the sample from top. This causes lower contrast for the lower parts of cell wall structure, as seen in Fig. 7.2. This step aims to obtain the highest contrast available without letting noise taint the details, and it is manually executed for a few reasons. Dynamic intensity range of each sample batch, and even each individual sample, is different. Furthermore, the noise profile for each image and the orientation of each guard cell pair according to the imaging plane vary. In order to obtain the best contrast difference possible, ImageJ enhance contrast tool is employed [181]. An inspection of the sample indicates the appropriate maximum and minimum intensity values.

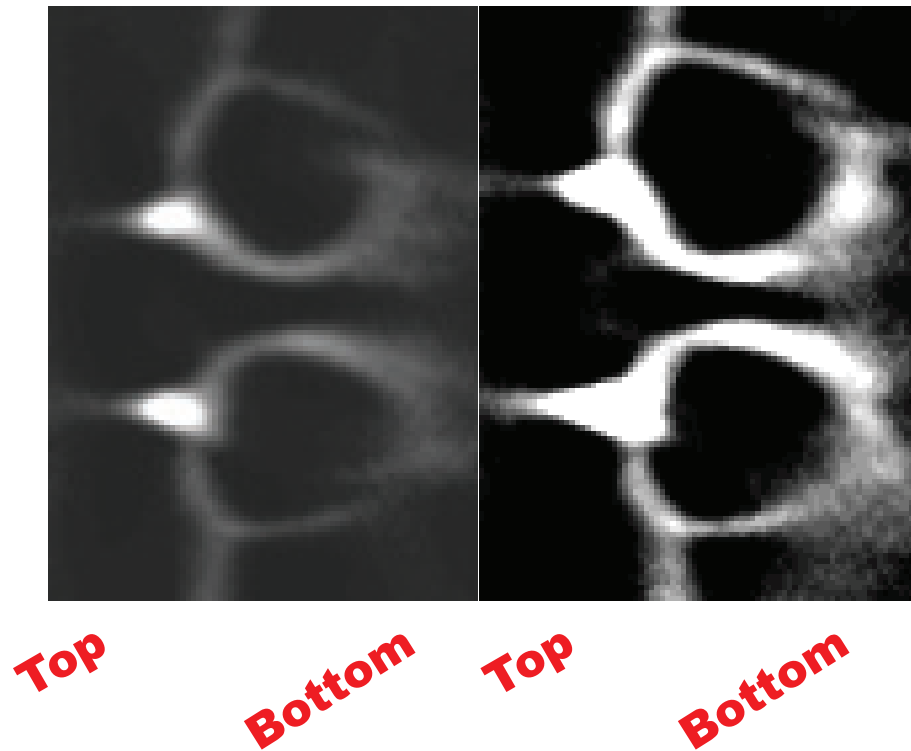


Figure 7.2: Contrast enhancement difference. The borders towards the bottom are indiscernible.

### 7.2.2 Initial Surface Extraction

Due to loss of fluorescent material deeper into the sample, some parts of the stomatal wall structure is either blurry or absent, breaking the visual continuity of the wall. Hence, methods utilized in the segmentation of the stomatal surface from confocal microscopy images should be robust to discontinuities in the surfaces. A method proposed by [182] puts forward a hybrid evolution framework for surfaces. This work improves the original work that is based on region based active contours algorithm introduced in [183]. As opposed to snakes or active contours which evolve according to edge information (please see Chapter 3), these hybrid active contours evolve according to region information. They have an initial sense of inside and outside regions of the evolving contour. As seen in their formulation Eq.(7.1), the step function (Heaviside) gives this property to the energy functional.

The original formulation of the functional follows as

$$\begin{aligned} \varepsilon(\phi, \mu_{in}, \mu_{out}) = & \lambda_1 \int_{\Omega} (I - \mu_{in})^2 H(\phi) d\Omega \\ & + \lambda_2 \int_{\Omega} (I - \mu_{in})^2 (1 - H(\phi)) d\Omega \\ & + \alpha \int_{\Omega} H(\phi) d\Omega + \beta \int_{\Omega} |\nabla H(\phi)| d\Omega, \end{aligned} \quad (7.1)$$

where  $\mu_{in} = \frac{\int_{\Omega} I H(\phi) d\Omega}{\int_{\Omega} H(\phi) d\Omega}$  and  $\mu_{out} = \frac{\int_{\Omega} I (1-H(\phi)) d\Omega}{\int_{\Omega} (1-H(\phi)) d\Omega}$  are average intensities inside and outside initial contour.  $H(\phi)$  is Heaviside function. The first two terms in Eq.(7.1) measure the variation inside and outside the active contour. Third and fourth terms respectively measure total area covered by the contour and the length of the contour.

In [182], they propose a hybrid energy functional that combines the region based properties and contour properties. The proposed functional is formulized as

$$\varepsilon(\phi) = -\alpha \int_{\Omega} (I - \mu) H(\phi) d\Omega + \beta \int_{\Omega} g(|\nabla I|) |\nabla H(\phi)| d\Omega, \quad (7.2)$$

where  $g(r)$  is a strictly decreasing function such as  $g = \frac{1}{1-c|\nabla I|^2}$  as mentioned in section 3.5.2, and  $\mu$  is the lower bound of the intensity of the target object. The first term enforces that contours contain regions with intensity higher than  $\mu$ . The second term is borrowed from geodesic active contour functional in Eq.(3.37). Fig. 7.3 demonstrates the process. It can be seen that pavement cell and other noise elements creep in the segmentation which requires a semi-automated involvement to clean up. Another problem with initial segmentation method is that larger gaps remain open after this step. The segmentation surface does not stop at the gap, causing a collapsed look for evolving surface, as seen in Fig. 7.4. In order to alleviate this problem, we do two passes for initial surface extraction.

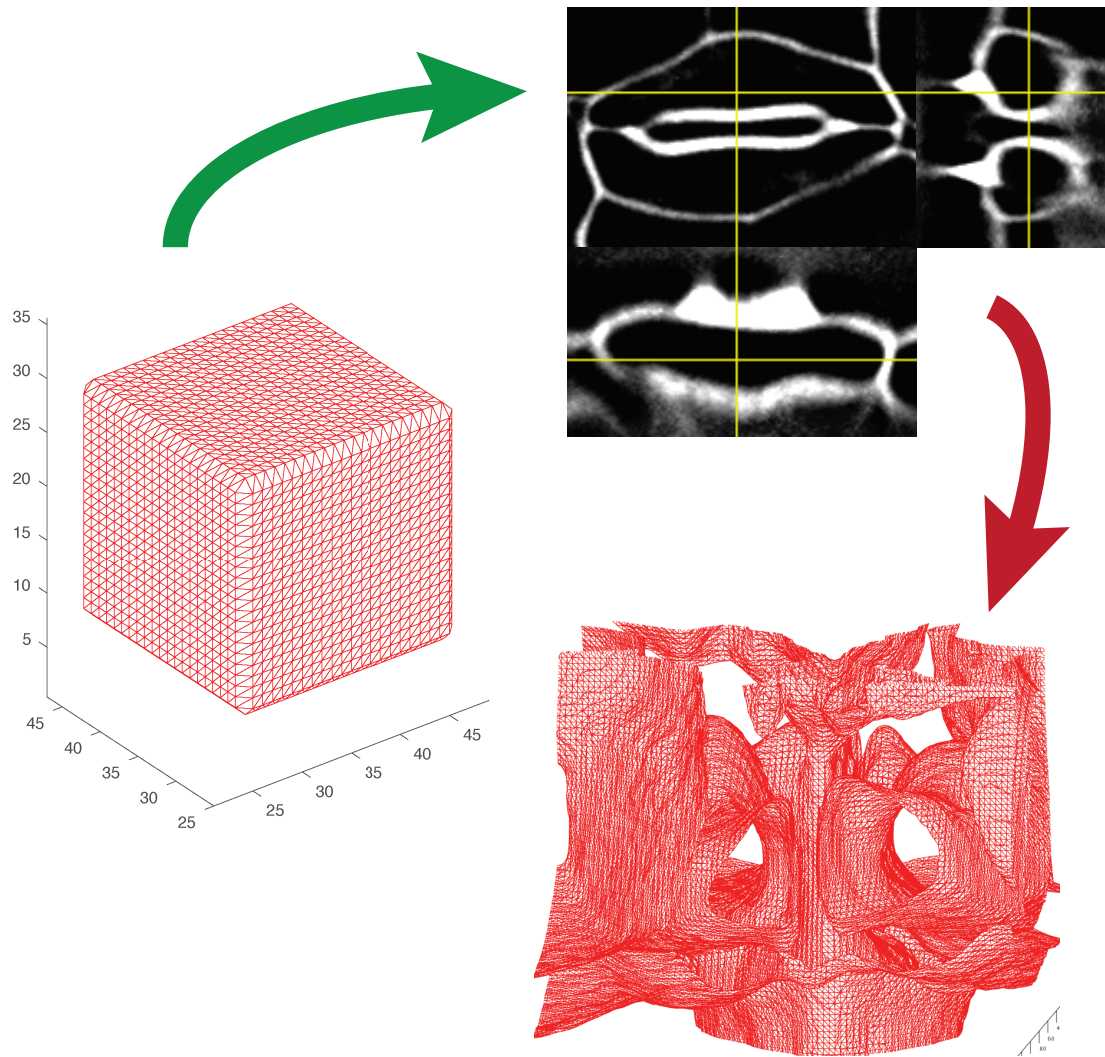


Figure 7.3: Initial surface extraction through 3D active surfaces. A cube that is the initial surface evolves according to guard cell shape leading to the initial surface segmentation. The second step improves this crude segmentation. The stoma guard cell pair is shown on XY (top-left), YZ (top-right), XZ planes at the yellow crosshair position.

### Two-Pass Surface Segmentation

In order to solve the issue with gaps, or low contrast areas, we devise a two-pass approach to the initial segmentation process. We first run the segmentation with lower intensity lower bound for the guard cells. This relaxes the constraints on the evolving surface to handle the low contrast gaps, but other low intensity level undesired structures get also included. In order to separate these structures, we run

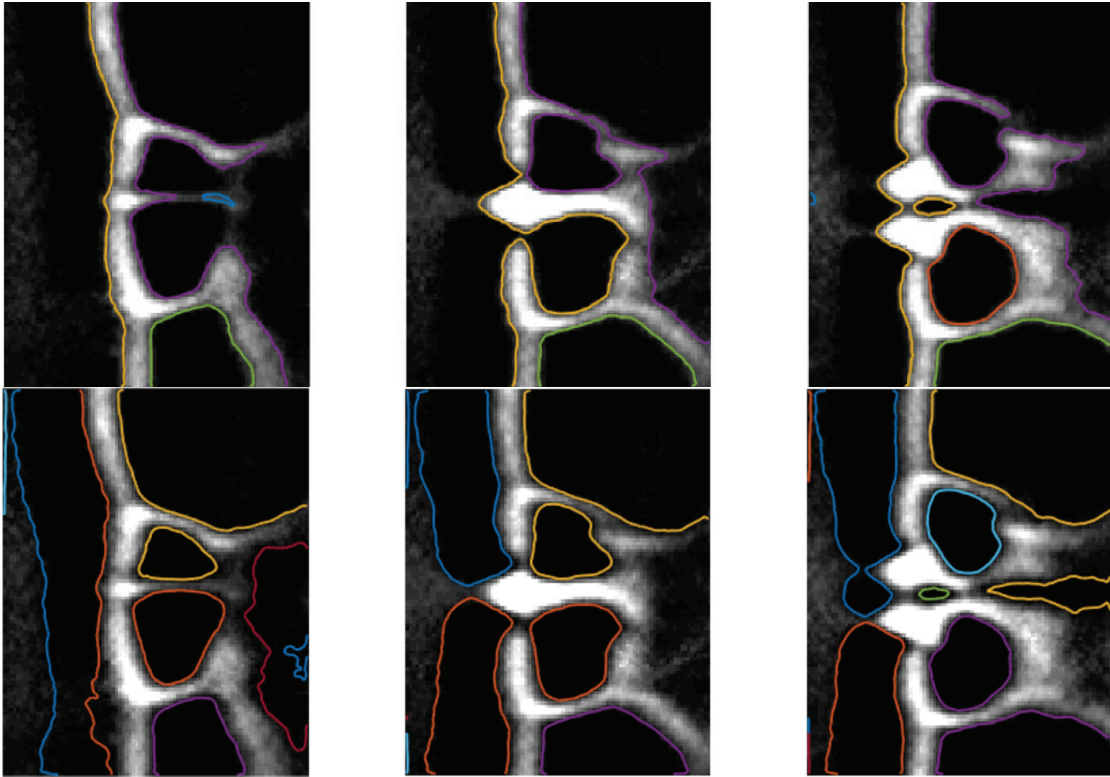


Figure 7.4: The low contrast areas cause segmentation errors as the segmentation surface collapse into the guard cell, as seen in first row. In order to get the whole wall system, we run the same segmentation with two times with lower (second row) and higher (first row)  $\mu$  values from Eq.(7.2). The lower  $\mu$  values causes the detection of pavement cells and noise, while higher  $\mu$  presents tighter fit with a probability of inward collapse.

the segmentation again with a higher intensity lower bound. This step eliminates noisy components. The first row of Fig. 7.4 shows results for a higher lower bound.

### 7.2.3 Inner and Outer Wall Separation

The second stage of the method involves closing bigger gaps, with the exception of pavement cells, that slipped through active surface-based segmentation and distinguishing between the inner and outer walls. As the individual walls of a stomata guard cell are cylindrical, the whole wall can be seen as a stack of 2D concentric circles. If the inside of a guard cell can be distinguished, then the inner and outer walls can also be detected. The circular shape of the walls can

be considered a heuristic. Large gaps point to use of another active contour-based approach. Hence, we propose using polar snakes on the segmentation surface from the initial segmentation.

This requires selecting two relative origin points in the middle of each guard cell (left and right). The origin points are intermittently required from the user as the guard cells are not perfect cylinders and the center of each cell changes from one tip to the other tip of a guard cell. The change in the position of the center is detected by intermittently obtaining origin points from the user. Surface edges show up as linear structures in the polar coordinate system  $(\rho, \Theta)$  for each image slice. The inner contour is the closest line to the  $\rho = 0$  line, followed by the outer contour. As stated above, these are rough lines with possible gaps. In order to extract them fully, we utilize linear snakes in the polar space. Snakes are line structures that expand or shrink and change shape according to a energy functional in order to segment structures in a 2D image. The energy function is often formulated as Eq.(3.36), however image is polar transformed in our case. The same image is repeated three times and concatenated vertically to handle periodic nature of  $\Theta \in [0, 2\pi]$ .

A snake is initialized at  $\rho = 0$  and evolved to fit onto the inner wall. Once the inner wall is obtained, the detected inner wall is taken as another initial snake and evolved again to fit the outer wall. This process is repeated for each slice to obtain the inner and outer walls. At the end of each evolution, the contour points are put through a Savitzky-Golay filter [184] for smooth contour results, and they are then transformed back into the Cartesian coordinate system, as seen in Fig. 7.5.

This approach requires a slight change at the junction regions of the guard cells because the outer wall is a single large contour. Snakes from left and right guard cells coincide, yielding erratic segmentation results. In order to solve this problem, a decision filter is added to check whether the outer wall exists between the two origin points. Unless the slice is at the junction region of the stomatal complex, the line from one origin point to another crosses four structures (inner-outer-outer-inner walls) for two cells. In the other case, it crosses two structures (inner-inner walls). The filter determines whether the slice is located at the junction region and observes the binary skeleton of the whole image.

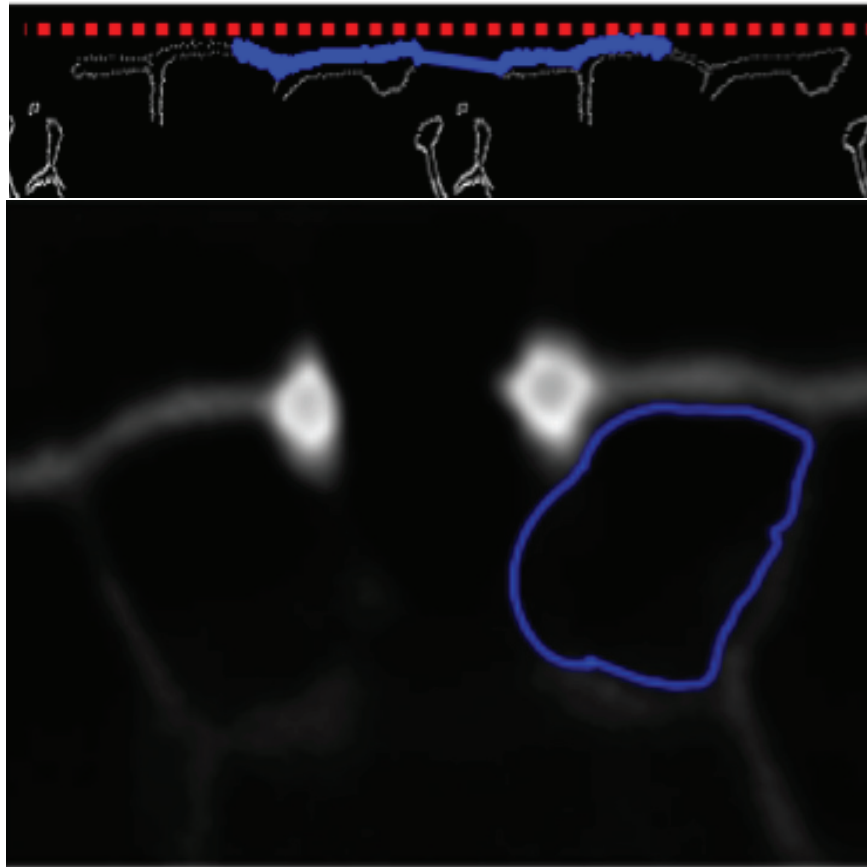


Figure 7.5: Top: Polar snake in polar coordinates ( $\rho$ ,  $\Theta$ ). Bottom: Polar snake transformed back to Cartesian coordinates. This process is conducted for each slice in volume of interest (VOI). Note the bulbous region on the inner side of the contour.

The skeleton part that extends from the top to the bottom of the junction region, including the junction wall, is added to the final segmentation.

It is important to bear in mind from our observations that the gaps due to contrast loss are sometimes too large, meaning that the polar snakes complete them as perfect lines in the polar coordinate system. When transformed back into the Cartesian coordinate system, these polar lines become perfect circles, which produces bulbous regions on the extracted surface, as demonstrated in Fig. 7.5. Such regions may require manual manipulation to obtain a good model.

## 7.2.4 Final Surface Smoothing

Extracted surfaces require a final 3D smoothing operation. This is because the polar snakes are initialized at each slice. Hence, the final result of each snake differs from one slice to another. This causes the final surface to look jittery and requires the application of an averaging filter to smooth out the surface. For this purpose, a cubic moving average filter on surface point positions is applied. This process also closes any gaps that persist if that location has enough neighboring points. The size of the cubic filter can be adjusted for different sample conditions. At the end of this stage, we have a rough 3D model of the stomatal complex as points.

## 7.2.5 Rendering and Manual Postprocessing

After the surface points are obtained, they need to be rendered in 3D and smoothed further for better visualization. An external rendering tool such as MeshLab can be leveraged for this purpose [185]. The bulbous regions can be removed and closed using the editing tools available in MeshLab. After the bulbous regions are removed, the surface is reconstructed. The reconstructed surface is smoothed for a second time to achieve better visualization, this time using Taubin surface smoothing [186]. Our observation suggest that straight Taubin smoothing step is enough for inner walls as they are not exposed to pavement cells and other noise artifacts. However, the outer walls require manual manipulation.

# 7.3 Evaluation

## 7.3.1 Data

Rosette leaves were collected from 3- to 4-week-old plants. Stomatal opening was induced by incubating leaves in a buffer containing 20 mM KCl, 1 mM CaCl<sub>2</sub>, and 5 mM MES-KOH, pH 6.15, in light for 2.5 h. Stomatal closure was induced by incubating leaves in a buffer containing 50 mM KCl, 0.1 mM CaCl<sub>2</sub>, and 10 mM MES-KOH, pH 6.15, in the dark for 2.5 h. Leaves were then stained with 100  $\mu$ g/ml propidium iodide (PI, Life Technologies; catalog no. P3566) for 5 min



before imaging. Z-stack images were collected on a Zeiss Axio Observer microscope attached to a Yokogawa CSU-X1 spinning disk head with a 63X 1.4 numerical aperture immersion oil objective, using a 561 nm excitation laser and a 617/73 nm emission filter with a step size of  $0.2 \mu\text{m}$ . Z-stack images were then subject to three-dimensional blind deconvolution using the AutoQuant X2 (Media Cybernetics) software.

### 7.3.2 Results

The candidate guard cell pairs are chosen from the samples. For the first sample, the initial segmentation is run at  $\mu = 20$  and  $\mu = 55$  after manual contrast enhancement. Then semi automated polar snake segmentation is run on the volume, followed by  $3 \times 3 \times 3$  cube neighborhood average smoothing. The initial segmentation is run at  $\mu = 52$  and  $\mu = 102$  for the other guard cell pair. The same average smoothing filter is employed. During the whole process, the polar snake step extracts inner left or right and outer left or right cells in one process. However, the manual manipulation part takes care of each guard cell. The whole model can be completed around one hour or one hour and a half. Fig. 7.6 displays the results of the each step of manual manipulation. As there is no ground truth for the data collected, we measure how realistically our model captures guard cell geometry by comparing it to other models of stomata. One such model was proposed in [12]. Fig. 7.6 also compares that model to our extracted 3D surface. While other model represents guard cell pair close to ellipse tubes, our model has bumps and deflations. A similar model to one proposed in [12] was employed in explanation of opening and closure of stomata in [13]. Fig. 7.6 also compares these models to our extracted model for closed and open cases. It can be seen that the deflated part for a closed stoma at the bottom of the figure is captured in our 3D model.

## 7.4 Discussion

One of the shortcomings of our method is the jitter from one slice to other slice. The randomness creeps in with initialization of snakes with each slice. Even though the averaging step takes care of some jitter, however there are some bulbous re-

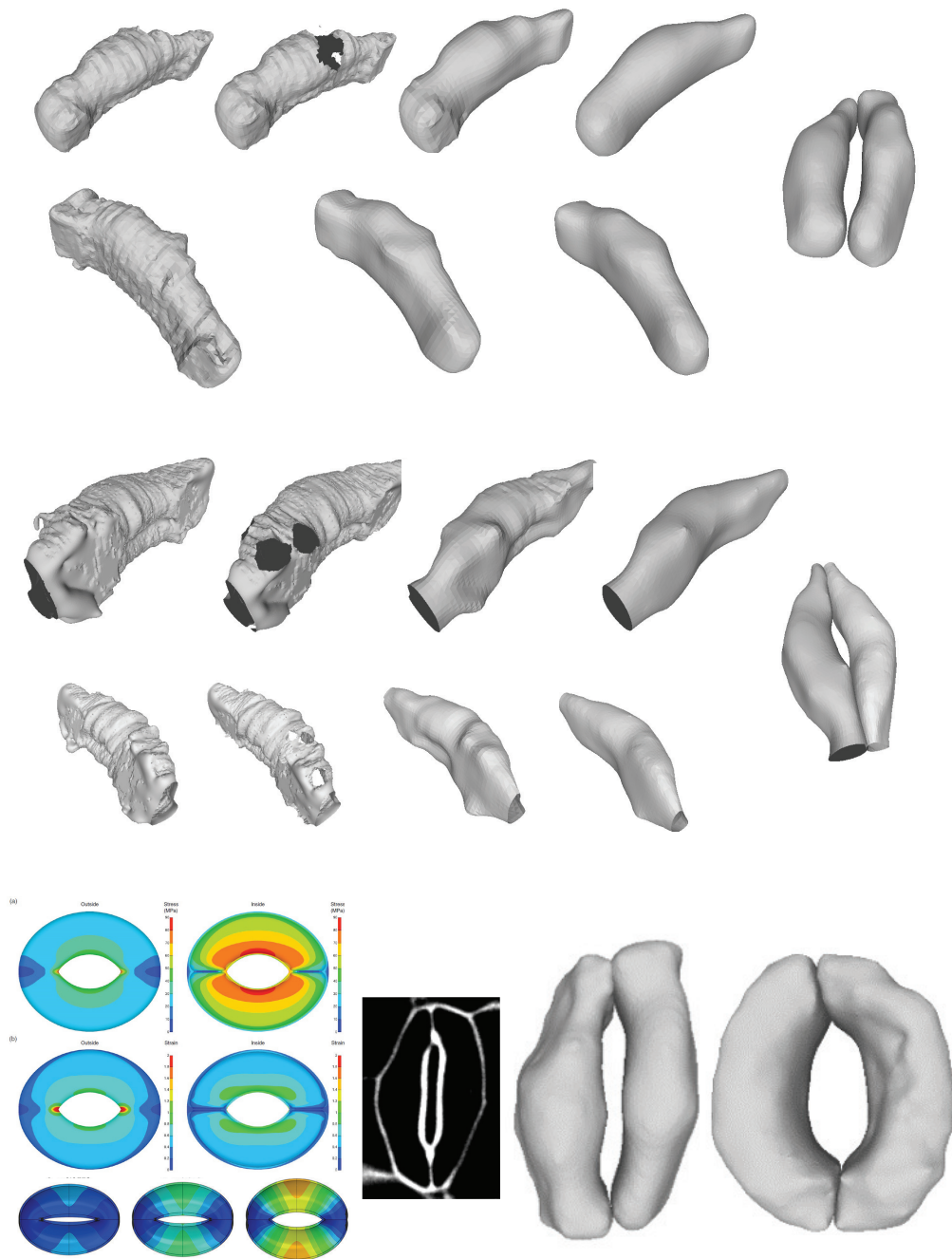


Figure 7.6: Results on the collected samples. The first four rows demonstrate the process of manual manipulation and smoothing after the surface points are obtained and resulting guard cell pair. Bottom row compares our results to other models that are recently employed in [12, 13]. The actual close stoma XY slice shows some deflations on sides which are captured by our model.

gions that are big enough to evade smoothing but too small for manual cropping. The polar segmentation should happen in 3D with one process. This might be included into the evolution information. Polar space can be replaced by cylindrical or spherical coordinate system.

In terms of automation, the involvement of users can be further decreased. The seed points that are received for polar transform origin reference can be further automated. Only two end points of a guard cell can be obtained from the user. Within the current scheme, a local neighborhood polar snake segmentation gives some surface information for neighboring slices. These neighboring slices can be utilized and the center of mass computed from inner wall binary image can act as an origin point. The precision of origin point location is more crucial for outer wall, and the desired location precision can be obtained by looking at the segmentation of inner walls with previous origin point, which is not that accurate; but inner wall segmentation is more robust to location of origin as the contour is closer.

Another part that can be further automatized is rendered surface reconstruction (VCG) and Taubin smoothing functions. As Meshlab is an open source system, these functions can be incorporated into our software package at the end of polar snake segmentation.

Two things that our system suffers is benchmarking and validation. A dataset with groundtruth markings are essential in measuring our performance and validating the improvements to the system. The stomata that were manually marked for landmarks in [132] can be passed to our pipeline, and the distance between our surface points and those points can be a measure of performance. The models that were previously proposed can also be run against the same data and we can compare the models in a more accurate manner.

The visual analysis of results show that our software package produces more realistic 3D surface models for different states of stomata. This might be useful for faster analysis in plant biology, or more accurate models in agricultural engineering. Measuring pore area, tangential length, guard cell size, surface area could be more reliable with this software.

## 7.5 Summary

In this chapter, we introduced a software package that can extract surface of stomatal complex more realistically compared to other models employed in previous plant biology studies. In our system, we extract initial surface using an active surface segmentation. This segmentation is improved with polar snake based inner/outer wall extraction through slices of the volume. Obtained surfaces are smoothed with 3D averaging filter. The surface points are rendered utilizing MeshLab program. Manual correction/cropping is followed by surface reconstruction and smoothing which provides final surface model. This software might be useful in plant biology and agricultural research.

## Conclusion and Future Work

In this chapter, we provide conclusions on the dissertation. The first section summarizes main themes of our research on demographic difference in aesthetic and affective response to visual features, visual balance, and 2D and 3D biomedical image processing. Following section presents suggestion for improvements to current work and possible future work. The chapter ends with discussing broader impacts of our results.

### 8.1 Summary

In this dissertation, we investigated how to make data more manageable from different perspectives. We considered content creation and analysis in our crowdsourcing and visual balance studies. Then, we investigated computational approaches to segmenting guard cells and microtubules.

For visual balance analysis, we challenged the general assumption in computational aesthetics that computational saliency can predict visual center of an image, hence its visual balance. We created a unique balance dataset through an online survey, extracted human, text, and vanishing point positions. It was shown that the positions of people and vanishing points in the pictures contributed to a better visual center prediction in a hierarchical linear regression framework. Continuing the investigation of visual features that are relevant to the computational aesthetics research, we attempted to understand how different people groups respond to these visual features aesthetically and emotionally. We collected a large number

of images aesthetically and emotionally rated via crowdsourcing. Through latent group analysis, we sought after different groups of people that demonstrate different responses to the same group of visual features.

In biomedical image processing field, we looked into tracing of tubular structures with inherent visual gaps through active curves and tensor voting. We showed that our proposed method performs better than the baseline method in a survey of plant biologists. In parallel, we utilized active contours and surfaces to analyze 3D stomatal guard cells. Both systems have the potential to save substantial time for scientists.

In conclusion, these results can be put into use in different systems that help people from different backgrounds catch up with the speed of data. A compositional assistant, a better personalized content manager might help people with little expert knowledge for content creation. Scientific analysis can be accelerated through our proposed systems.

## 8.2 Areas of Future Work

The results detailed in previous chapters have implications for future research. We discuss these implications and the possible implementation of our findings in the design of future systems.

### 8.2.1 Improvements to Balance Study

As we have discussed, visual balance can be an effective way to achieve striking photography. The study that we conducted can be improved and put into a retrieval system.

**Visual Weight.** As demonstrated in Chapters 2 and 4, compositional balance can be achieved through the use of different visual elements. Any attempt to achieve it should take these visual features into account. Visual saliency was shown to be insufficient for visual center representation, and hence compositional balance. The results demonstrate that we need to consider higher-level features that provide more information to the viewer.

The task does not rely only on the calculation of these features and the incorporation of them into a vectoral framework, however. One of the challenges that makes this problem even more interesting is that we don't know how these visual features interact with each other to achieve compositional balance. The association of visual information that is extracted from images with the perception of people is necessary for this line of research. This may require a machine learning approach where the ground-truth is obtained from the user study defined in Chapter 4. Machine learning can be employed in two different ways. First, a model that is learned from the data can be utilized to directly predict a visual center score for an image. In another approach, a classification system can categorize an image into four groups that were in our balance study. These results can be considered as a high-level feature that can be employed.

**Extended User Study.** Although books provide art theories about balance, our preliminary user study showed that this concept is hard to understand. However, there are a few instances on which people agree. The preliminary study indicated that the dynamic balance and symmetry are sometimes confused. There are cases in which individuals agree on imbalancedness but attribute it to different visual elements or choose opposite extremes for slider positions. The preliminary study is helpful in understanding the difficulties involved in conducting a user study on compositional balance. Once the interface is established, the data can be collected in two different ways:

- The first way is to reach out to many people of different ages and backgrounds through crowdsourcing and investigate how they understand the concept of balance after they are exposed to the tutorial provided before the study. The online crowdsourcing platform Amazon Mechanical Turk can be employed for this task.
- Another way is to employ the same interface but to choose users who are photography experts to benefit from their opinions and knowledge.

## 8.2.2 Integration of Balance into Composition Retrieval

Once the concept of balance has been investigated, different possibilities present themselves. The system can be used to provide a predicted balance point position

for a query image. With this value, two potential problems can be tackled. First, in line with the psychology studies presented in Section 2.2, the image may be manipulated through gamma modification or cropping (retargeting) in order to achieve balance if the image is off-balance. The ideal system would give real-time feedback to the user by indicating the elements that are causing imbalance. As seen in Fig. 8.1<sup>1</sup>, the user can be guided while taking a picture.

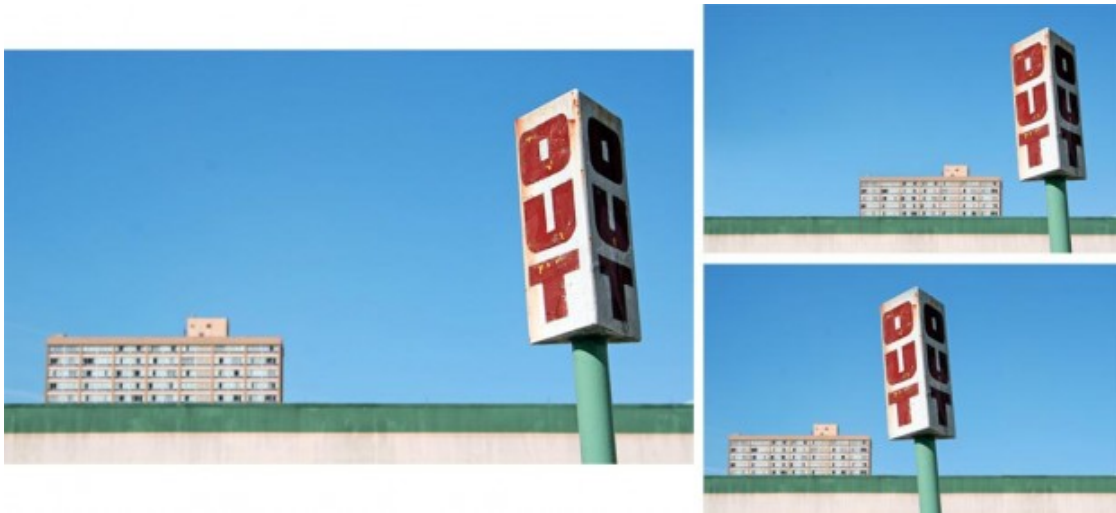


Figure 8.1: Using the calculated balance point position, real-time feedback can be given to users [14].

The other main problem that can be addressed is the retrieval of similar composition images. Once the distribution of visual weight over an image has been obtained, the similarities between the images can be quantified. As compositional balance is highly related to spatial distribution, the images can be divided into grids and then the average feature vector can be computed for each individual grid area. For two different images, the average vectors of the areas can be compared through a similarity measure such as cosine distance.

### 8.2.3 Personalization and Crowdsourcing Survey Design

The results of the study about demographic differences in emotions and aesthetic perception due to visual features offer a chance to modify content according to viewers' taste. Depending on the ethnicity, income, and education level of a user,

<sup>1</sup>SLR Lounge Photography 101: Composition



online advertisements can be manipulated. As the study in Chapter 5 shows, simplicity or the use of light can create different impressions for different demographic groups. For a user who belongs to a specific demographic groups, brighter and simpler images can be provided to elicit a positive response to the content.

On the other hand, survey designers may need to take these differences into account while they are recruiting workers from AMT and designing their surveys. They may need to stratify the workers that take their surveys in order to have a balanced participant base.

### 8.2.4 Stomatal Research

With the methods detailed in Chapter 6 and 7, we can pinpoint the location of the microtubules and inner-outer walls of a stoma guard cell pair. This work can be improved further and utilized in various aspects of plant biology research.

**More Groundtruth Data For Benchmarking.** The methods proposed could be improved with availability of more groundtruth data. The knowledge of experts can be leveraged through an annotation effort. We can have experts trace microtubules through an annotation software. The average distance between groundtruth curves and computed curves can be measured and reported.

**Microtubule Detection.** Future work includes extending our approach to 3D volumes. A formulation of tensor voting in higher dimensions is available, but it has not been implemented. One possible improvement could be the integration and implementation of 3D tensor voting with the 3D volume SOAX framework. Another option is to keep the core algorithm for 2D as it is, and save the slice indices for each pixel in the memory while obtaining the maximum intensity projection images. At the end of tubular structure detection, a detected pixel can be put back into 3D volume according to its slice index. This proposed framework could be added to the end of a learning pipeline to refine segmentation results. As more labeled data become available, learning-based methods can be better leveraged for segmentation and tracing problems. For example, in the future, we could utilize deep learning to learn more about abundant visual representations from big data created by experts [179, 180].

These features can be directly used in an end-to-end deep learning framework or can be fed into another machine learning algorithm such as the support vector machine (SVM) to obtain the improved segmentation results for further analysis.

**Segmentation of CESAs.** Unlike microtubules, CESAs are blobular structures. The eigen analysis for a Hessian image could be applied to the segmentation of blob structures by slightly modifying the the vesselness equations. The size of the eigens could be compared in a similar manner to tubule detection.

A scheme such as the one proposed in [187] could be put to use to detect blobular structures, since we are using 2D structures for the validation of our arguments. In the case of eigenvalues where  $|\lambda_1| \leq |\lambda_2|$ , the blobness can be formulated as

$$B_\sigma(x, y) = \begin{cases} \frac{|\lambda_1|}{|\lambda_2|} e^{-R_c^2/2}, & \text{if } \lambda_i < 0 \text{ for } i = 1, 2, \\ 0 & \text{otherwise.} \end{cases} \quad (8.1)$$

where  $R_c = k - 2\sigma$ .  $\sigma$  is the scale of the Gaussian smoothing filter, while  $k$  is called the *local object scale* and defined as “the radius of the largest hyperball centered at the pixel such that all pixels within the ball satisfy a predefined intensity homogeneity criterion.” This can be considered as helpful information about the local structure’s size. In our case, average CESA sizes are known.

**Colocalization Analysis.** Once the structures are segmented, their localization can be analyzed by measuring their overlap. One way to do this is by overlaying a 2D Gaussian distribution on the segmented MTs and segmented CESAs. The overlap of these Gaussians can be considered as computing a measure of colocalization.

**3D Modelling of Stomata.** The surface extraction scheme that was proposed in Chapter 7, presents room for improvement. The full automation of the process could reveal the true potential of this approach. The extracted surfaces could be put to use to simulate stoma behavior more accurately. Accuracy would provide us with the ability to predict and manipulate stoma behavior. The following section provides a broader impact of our results.

### 8.3 Anticipated Outcomes and Implications

Our results regarding differences in the affective and aesthetic responses of different latent demographic groups can be employed for different research purposes. One immediate use case may be personalization in the user experience. As stated above, individuals interact with various interfaces including Web pages and social media platforms. Their user experiences can be improved adapting these interfaces according to their demographic background. Our approach to the topic of visual balance can be considered a stepping stone to other exploratory studies on aesthetic principles such as rhythm and unity. Our approach can serve as an example of how other features can be quantified.

Our proposed system for the computerized analysis of stoma might lead to faster processing while retaining accuracy in hypothesis testing for scientists. Extracted 3D surfaces can lead to better physical modeling, which gives way to better understanding of stomatal behavior. Hence, we can control stoma behavior. The ability to control stoma behavior has implications in terms of agricultural irrigation planning. This can provide great benefits in our struggle against global warming.

## Secondary Research For Stomatal Analysis

This section includes various work done for secondary and small projects to help with analysis of stomata and their effect on physical appearance of the plants. One work helped to analyze the effect of a gene (Polygalacturonase) on seedling growth development along with stomatal functions [188]. The other helped with analysis of pore area, guard cell tangential length and size calculation of stoma guard cells [132].

### A.1 Rosette Area Segmentation

Seedlings on plates were scanned on a Scanjet 8300 scanner (HP) at 600 dpi. Root length and hypocotyl length were measured in ImageJ (W.S. Rasband, U. S. National Institutes of Health, Bethesda, MD; <https://imagej.nih.gov/ij/>). Relative growth rate of roots was calculated on an individual basis. Rosette images were taken with a Nikon D5100 DSLR camera and were computationally segmented. Raw rosette images were in the RGB color space, and were converted to the HSV (hue, saturation, and value) color space using the OpenCV image processing software library. A single range of green colors was chosen from the histogram of hue values to separate the foreground rosette from the background. Threshold cutoffs were applied to obtain green areas with elevated saturation values according to this range. The binary image masks were morphologically processed to fill holes



Figure A.1: The segmentation of leaves by thresholding according to their hue, saturation and value(brightness).

and remove small artifacts in the background. The foreground regions delineated by the mask were selected from the original image.

## A.2 Stoma Geometrical Properties Analysis

A semi-computerized active contours-based method was adopted to segment and measure the pore and cell-pair areas [189]. In this approach, the user initializes a closed curve by entering a few points. The initialized curve expands and evolves according the edge map of the image, fitting to the irregularly shaped object. The evolution is driven by an optimization scheme where the cost is minimized if the curve fits the edges smoothly. Pore area values across each z-stack were obtained via aforementioned method. To account for outliers due to contrast issues, a robust nonlinear regression (fourth-order polynomial) was applied on the area profile graph, where the variance was assumed to have a Cauchy distribution [190]. The minimum of the regression curve was picked as the pore area in number of pixels (Figure A.2). As the physical area of a pixel in the XY plane is  $0.0409 \mu\text{m}^2$ , the pore area in pixels was multiplied by this number to obtain the actual area. A connected component analysis was carried out for minimum pore area segmenta-

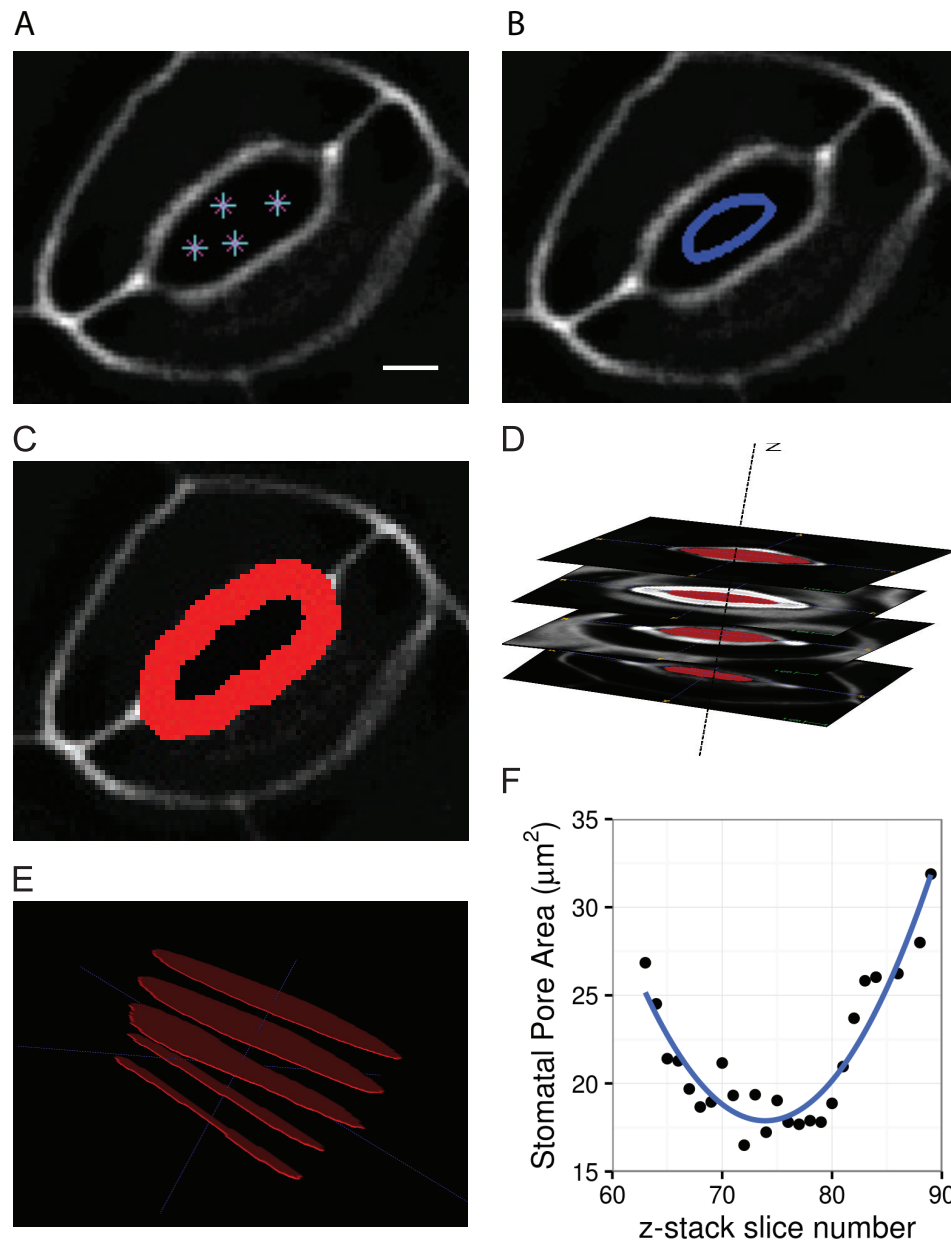


Figure A.2: Segmentation of pore area in 3D through semi-automated 2D snakes. The pore area is segmented using snakes algorithm through an initial curve across slices (C-D). The distribution of pore areas is first cleaned from outliers, then a robust curve is fitted, and minimum is detected (E-F).

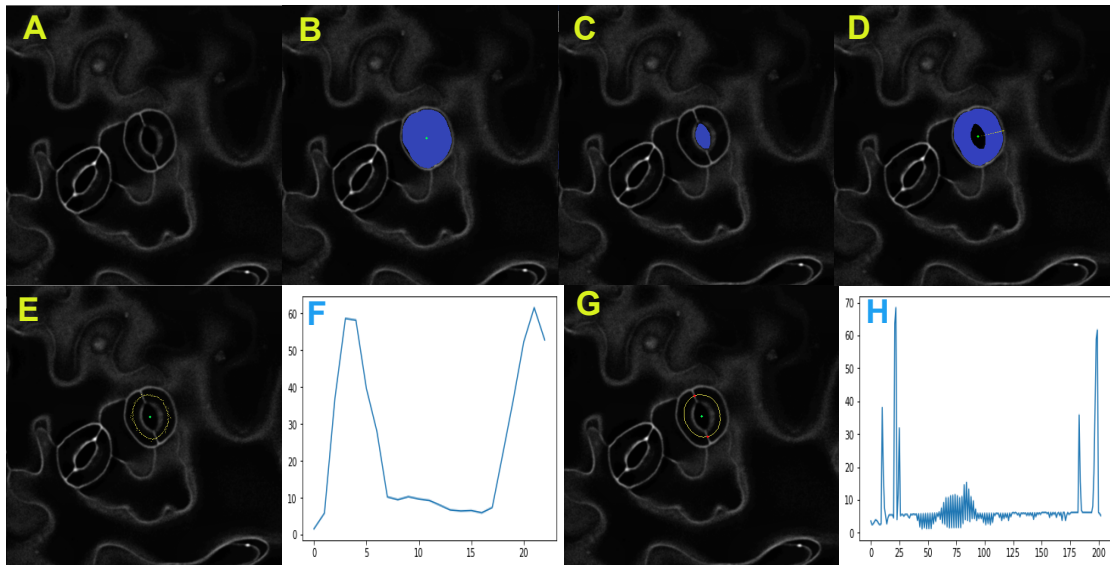


Figure A.3: Single cell including pore area is detected in a similar fashion to pore area segmentation, and the associated pore segmentation is subtracted to obtain guard cell (B-C-D). The junction point locations are determined via tangential tracing of intensity values (E-F). Radial tracing of intensity is conducted to detect cell walls, hence to find the medial tangential axes (G-H).

tion. The major and minor axes of the connected component were computed to measure the height and the width of the pore area, respectively. A similar analysis was administered for cell geometry measurements. Active contours segmentation was carried out to obtain the image slice that contained the minimum pore area, or another slice within four-slice neighborhood, where guard cell pair perimeter contrast was better. The same connected component analysis was conducted to measure stomatal complex length and width. After getting binary masks for cell and pore area, the tangential length of each cell was computed. For this task, intensity profile analysis in radial direction originating from the center of mass for the pore opening was conducted at each angle within the interval  $[0^\circ, 359^\circ]$ . The intensity local maxima representing pore and cell walls are detected and the midpoint of the line segment between the two points was calculated. This method provided a medial axis for each cell within the pair.

The medial axis was smoothed out through the Savitzky-Golay filter, where subsets of data were re-fit via polynomial least squares [184]. The medial axis curves were simplified through polynomial approximation to dispose of unnecessary meandering and jitter (Figure A.3). The arc length of the simplified medial axes was computed to measure the tangential length of individual cells.



# Bibliography

- [1] LLOYD, C. and J. CHAN (2004) “Microtubules and the shape of plants to come,” *Nature Reviews Molecular Cell Biology*, **5**(1), pp. 13–23.
- [2] ARNHEIM, R. (1954) *Art and Visual Perception: A Psychology of the Creative Eye*, Univ. of California Press.
- [3] MCMANUS, I., K. STÖVER, and D. KIM (2011) “Arnheim’s Gestalt theory of visual balance: Examining the compositional structure of art photographs and abstract images,” *i-Perception*, **2**(6), pp. 615–647.
- [4] KAHRS, J., S. CALAHAN, D. CARSON, and S. POSTER (1996) “Pixel cinematography: a lighting approach for computer graphics,” *ACM SIGGRAPH Course Notes*, pp. 433–42.
- [5] RUSSELL, J. A. (1980) “A circumplex model of affect,” *Journal of Personality and Social Psychology*, **39**(6), pp. 1161–1178.
- [6] LU, X. (2016) *Visual Characteristics for Computational Prediction of Aesthetics and Evoked Emotions*, Ph.D. thesis, The Pennsylvania State University, College of Information Sciences and Technology.
- [7] GONZALEZ, R. C. and R. E. WOODS (2007) *Digital Image Processing*, 3 ed., Prentice Hall.
- [8] ROTTENFUSSER, R., E. E. WILSON, and M. W. DAVIDSON (2016), “Basic Microscopy - Point Spread Function,” .  
URL <http://zeiss-campus.magnet.fsu.edu/articles/basics/psf.html>
- [9] WEICKERT, J. (1999) “Coherence-Enhancing Diffusion Filtering,” *International Journal of Computer Vision*, **31**(2-3), pp. 111–127.

- [10] MAGGIORI, E., H. L. MANTEROLA, and M. DEL FRESNO (2014) “Perceptual grouping by tensor voting: a comparative survey of recent approaches,” *IET Computer Vision*, **9**(2), pp. 259–277.
- [11] LANG, P., M. BRADLEY, and B. CUTHBERT (1997) *International affective picture system (IAPS): Affective ratings of pictures and instruction manual*, Tech. rep.
- [12] WOOLFENDEN, H. C., G. BOURDAIS, M. KOPISCHKE, E. MIEDES, A. MOLINA, S. ROBATZEK, and R. J. MORRIS (2017) “A computational approach for inferring the cell wall properties that govern guard cell dynamics,” *The Plant Journal*, **92**(1), pp. 5–18.
- [13] MAROM, Z., I. SHTEIN, and B. BAR-ON (2017) “Stomatal opening: the role of cell-wall mechanical anisotropy and its analytical relations to the bio-composite characteristics,” *Frontiers in Plant Science*, **8**, p. 2061.
- [14] LIN, C. (2013), “Understanding Balance In Photography,” .  
URL <https://www.slrlounge.com/understanding-balance-in-photography/>
- [15] MOORE, G. E. (2006) “Cramming more components onto integrated circuits, Reprinted from Electronics, volume 38, number 8, April 19, 1965, pp.114 ff.” *IEEE Solid-State Circuits Society Newsletter*, **11**(3), pp. 33–35.
- [16] DE MAURO, A., M. GRECO, and M. GRIMALDI (2015) “What is big data? A consensual definition and a review of key research topics,” *AIP Conference Proceedings*, **1644**(1), pp. 97–104.
- [17] YAO, L., P. SURYANARAYAN, M. QIAO, J. Z. WANG, and J. LI (2012) “Oscar: On-site composition and aesthetics feedback through exemplars for photographers,” *International Journal of Computer Vision*, **96**(3), pp. 353–383.
- [18] LAUER, D. and S. PENTAK (2011) *Design Basics*, 8th ed., Wadsworth Cengage Learning.
- [19] DATTA, R., D. JOSHI, J. LI, and J. Z. WANG (2006) “Studying aesthetics in photographic images using a computational approach,” in *Proceedings of the European Conference on Computer Vision*, vol. 3953 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 288–301.
- [20] JAHANIAN, A., J. LIU, Q. LIN, D. TRETTER, E. O’BIEN-STRAIN, S. C. LEE, N. LYONS, and J. ALLEBACH (2013) “Recommendation system for automatic design of magazine covers,” in *Proceedings of International Conference on Intelligent User Interfaces*, IUI ’13, ACM, pp. 95–106.

- [21] BHATTACHARYA, S., R. SUKTHANKAR, and M. SHAH (2010) “A framework for photo-quality assessment and enhancement based on visual aesthetics,” in *Proceedings of The International Conference on Multimedia*, ACM, pp. 271–280.
- [22] LIU, L., R. CHEN, L. WOLF, and D. COHEN-OR (2010) “Optimizing photo composition,” in *Proceedings of Computer Graphics Forum*, vol. 29, pp. 469–478.
- [23] MCMANUS, I., D. EDMONDSON, and J. RODGER (1985) “Balance in pictures,” *British Journal of Psychology*, **76**(3), pp. 311–324.
- [24] BRABHAM, D. C. (2013) *Crowd Sourcing*, MIT Press.
- [25] AROLAS, E. and F. GUEVARA (2012) “Towards an integrated crowdsourcing definition,” *Journal of Information Science*, **38**(2), pp. 189–200.
- [26] MURRAY, N., L. MARCHESOTTI, F. PERRONNIN, and F. MEYLAN (2012) “Learning to rank images using semantic and aesthetic labels.” in *British Machine Vision Conference*, pp. 1–10.
- [27] YE, J., J. LI, M. G. NEWMAN, R. B. ADAMS, and J. Z. WANG (2017) “Probabilistic multigraph modeling for improving the quality of crowd-sourced affective data,” *IEEE Transactions on Affective Computing*, **10**(1), pp. 115–128.
- [28] MARGE, M., S. BANERJEE, and A. I. RUDNICKY (2010) “Using the Amazon Mechanical Turk for transcription of spoken language,” in *Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing*, IEEE, pp. 5270–5273.
- [29] SOLEYMANI, M. and M. LARSON (2010) “Crowdsourcing for affective annotation of video: Development of a viewer-reported boredom corpus,” in *Proceedings of the ACM SIGIR Workshop on Crowdsourcing For Search Evaluation*, p. 4–8.
- [30] SMITH, S. M., C. A. ROSTER, L. L. GOLDEN, and G. S. ALBAUM (2016) “A multi-group analysis of online survey respondent data quality: Comparing a regular USA consumer panel to MTurk samples,” *Journal of Business Research*, **69**(8), pp. 3139–3148.
- [31] DIFALLAH, D. E., E. FILATOVA, and P. G. IPEIROTIS (2018) “Demographics and dynamics of Mechanical Turk workers,” in *Proceedings of the International Conference on Web Search and Data Mining*, ACM.

- [32] JOHANNES, L., R. G. PARTON, P. BASSEREAU, and S. MAYOR (2015) “Building endocytic pits without clathrin,” *Nature Reviews Molecular Cell Biology*, **16**(5), p. 311.
- [33] LI, S., L. LEI, C. R. SOMERVILLE, and Y. GU (2012) “Cellulose synthase interactive protein 1 (CSI1) links microtubules and cellulose synthase complexes,” *Proceedings of the National Academy of Sciences*, **109**(1), pp. 185–190.
- [34] COSGROVE, D. J. (2005) “Growth of the plant cell wall,” *Nature Reviews Molecular Cell Biology*, **6**(11), pp. 850–861.
- [35] BRINGMANN, M., E. LI, A. SAMPATHKUMAR, T. KOCABEK, M.-T. HAUSER, and S. PERSSON (2012) “POM-POM2/cellulose synthase interacting1 is essential for the functional association of cellulose synthase and microtubules in Arabidopsis,” *The Plant Cell*, **24**(1), pp. 163–177.
- [36] PAREDEZ, A. R., C. R. SOMERVILLE, and D. W. EHRHARDT (2006) “Visualization of cellulose synthase demonstrates functional association with microtubules,” *Science*, **312**(5779), pp. 1491–1495.
- [37] BOLTE, S. and F. CORDELIÈRES (2006) “A guided tour into subcellular colocalization analysis in light microscopy,” *Journal of Microscopy*, **224**(3), pp. 213–232.
- [38] FRENCH, A. P., S. MILLS, R. SWARUP, M. J. BENNETT, and T. P. PRIDMORE (2008) “Colocalization of fluorescent markers in confocal microscope images of plant cells,” *Nature Protocols*, **3**(4), pp. 619–628.
- [39] RIZK, A., G. PAUL, P. INCARDONA, M. BUGARSKI, M. MANSOURI, A. NIEMANN, U. ZIEGLER, P. BERGER, and I. F. SBALZARINI (2014) “Segmentation and quantification of subcellular structures in fluorescence microscopy images using Squash,” *Nature Protocols*, **9**(3), pp. 586–596.
- [40] FRANGI, A. F., W. J. NIESSEN, K. L. VINCKEN, and M. A. VIERGEVER (1998) “Multiscale vessel enhancement filtering,” in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, pp. 130–137.
- [41] RAJPUT, Y. M., R. R. MANZA, R. D. DEEPALI, M. B. PATWARI, M. SASWADE, and N. DESHPANDE (2016) “Design New Biorthogonal Wavelet Filter for Extraction of Blood Vessels and Calculate the Statistical Features,” in *Information Systems Design and Intelligent Applications*, Springer, pp. 647–655.

- [42] AZZOPARDI, G., N. STRISCIUGLIO, M. VENTO, and N. PETKOV (2015) “Trainable COSFIRE filters for vessel delineation with application to retinal images,” *Medical Image Analysis*, **19**(1), pp. 46–57.
- [43] HENEGHAN, C., J. FLYNN, M. O’KEEFE, and M. CAHILL (2002) “Characterization of changes in blood vessel width and tortuosity in retinopathy of prematurity using image analysis,” *Medical Image Analysis*, **6**(4), pp. 407–429.
- [44] BHADAURIA, H., S. BISHT, and A. SINGH (2013) “Vessels extraction from retinal images,” *IOSR Journal of Electronics and Communication Engineering*, pp. 2278–2834.
- [45] ZHU, W.-B., B. LI, L.-F. TIAN, X.-X. LI, and Q.-L. CHEN (2015) “Topology adaptive vessel network skeleton extraction with novel medialness measuring function,” *Computers In Biology and Medicine*, **64**, pp. 40–61.
- [46] XU, T., D. VAVYLONIS, and X. HUANG (2014) “3D actin network centerline extraction with multiple active contours,” *Medical Image Analysis*, **18**(2), pp. 272–284.
- [47] CHEN, L., M. MOSSA-BASHA, N. BALU, G. CANTON, J. SUN, K. PIMENTEL, T. S. HATSUKAMI, J.-N. HWANG, and C. YUAN (2018) “Development of a quantitative intracranial vascular features extraction tool on 3 D MRA using semiautomated open-curve active contour vessel tracing,” *Magnetic Resonance in Medicine*, **79**(6), pp. 3229–3238.
- [48] RUI, Y. and C. T. ANDERSON (2016) “Functional analysis of cellulose and xyloglucan in the walls of stomatal guard cells of Arabidopsis,” *Plant Physiology*, **170**(3), pp. 1398–1419.
- [49] WANG, M., K. YANG, X.-S. HUA, and H.-J. ZHANG (2010) “Towards a relevant and diverse search of social images,” *IEEE Transactions on Multimedia*, **12**(8), pp. 829–842.
- [50] WU, O., Y. CHEN, B. LI, and W. HU (2011) “Evaluating the visual quality of web pages using a computational aesthetic approach,” in *Proceedings of the International Conference on Web Search and Data Mining*, ACM, pp. 337–346.
- [51] ROSS, D. (1907) *A Theory of Pure Design: Harmony, Balance, Rhythm*, Houghton, Mifflin.
- [52] SAMUEL, F. and D. KERZEL (2013) “Judging whether it is aesthetic: Does equilibrium compensate for the lack of symmetry?” *i-Perception*, **4**(1), pp. 57–77.

- [53] LEYSSEN, M., S. LINSEN, J. SAMMARTINO, and S. PALMER (2012) “Aesthetic preference for spatial composition in multiobject pictures,” *i-Perception*, **3**(1), pp. 25–49.
- [54] LOCHER, P., S. GRAY, and C. NODINE (1996) “The structural framework of pictorial balance,” *Perception*, **25**(12), pp. 1419–1436.
- [55] LOCHER, P. J., P. J. STAPPERS, and K. OVERBEEKE (1999) “An empirical evaluation of the visual rightness theory of pictorial composition,” *Acta Psychologica*, **103**(3), pp. 261–280.
- [56] GUAN, S.-S. (2002) “A study of color harmony relating with area ratio,” in *Proceedings of the 9th Congress of the International Colour Association*, pp. 199–202.
- [57] LOCHER, P. J., P. JAN STAPPERS, and K. OVERBEEKE (1998) “The role of balance as an organizing design principle underlying adults’ compositional strategies for creating visual displays,” *Acta Psychologica*, **99**(2), pp. 141–161.
- [58] FREEMAN, M. (2007) *The Photographer’s Eye: Composition and Design for Better Digital Photos*, Taylor & Francis US.
- [59] JOSHI, D., R. DATTA, E. FEDOROVSKAYA, Q.-T. LUONG, J. Z. WANG, J. LI, and J. LUO (2011) “Aesthetics and emotions in images,” *IEEE Signal Processing Magazine*, **28**(5), pp. 94–115.
- [60] BATTIATO, S., M. MOLTISANTI, F. RAVÌ, A. BRUNA, and F. NACCARI (2013) “Aesthetic scoring of digital portraits for consumer applications,” in *IS&T/SPIE Electronic Imaging*, International Society for Optics and Photonics, pp. 866008–866008.
- [61] ZHENG, X. S., I. CHAKRABORTY, J. J.-W. LIN, and R. RAUSCHENBERGER (2009) “Correlating low-level image statistics with users-rapid aesthetic and affective judgments of web pages,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, ACM, pp. 1–10.
- [62] REINECKE, K., T. YEH, L. MIRATRIX, R. MARDIKO, Y. ZHAO, J. LIU, and K. Z. GAJOS (2013) “Predicting users’ first impressions of website aesthetics with a quantification of perceived visual complexity and colorfulness,” in *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, ACM, pp. 2049–2058.
- [63] ZHAO, S., Y. GAO, X. JIANG, H. YAO, T.-S. CHUA, and X. SUN (2014) “Exploring principles-of-art features for image emotion recognition,” in *Proceedings of The International Conference on Multimedia*, ACM, pp. 47–56.

- [64] LOY, G. and A. ZELINSKY (2003) “Fast radial symmetry for detecting points of interest,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **25**(8), pp. 959–973.
- [65] LOY, G. and J.-O. EKLUNDH (2006) “Detecting symmetry and symmetric constellations of features,” in *Proceedings of the European Conference on Computer Vision*, Springer, pp. 508–521.
- [66] THUMFART, S., R. JACOBS, K. V. HAAK, F. W. CORNELISSEN, J. SCHARINGER, and C. EITZINGER (2008) “Feature based prediction of perceived and aesthetic properties of visual textures,” *Proceedings of Materials & Sensations*, pp. 22–24.
- [67] MACHAJDIK, J. and A. HANBURY (2010) “Affective image classification using features inspired by psychology and art theory,” in *Proceedings of The International Conference on Multimedia*, ACM, pp. 83–92.
- [68] TAMURA, H., S. MORI, and T. YAMAWAKI (1978) “Textural features corresponding to visual perception,” *IEEE Transactions on Systems, Man and Cybernetics*, **8**(6), pp. 460–473.
- [69] ARBELAEZ, P., M. MAIRE, C. FOWLKES, and J. MALIK (2011) “Contour detection and hierarchical image segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **33**(5), pp. 898–916.
- [70] VAN DE WEIJER, J., C. SCHMID, and J. VERBEEK (2007) “Learning color names from real-world images,” in *Proceedings of the Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 1–8.
- [71] MOJSILOVIC, A. (2005) “A computational model for color naming and describing color composition of images,” *IEEE Transactions on Image Processing*, **14**(5), pp. 690–699.
- [72] MORRISS, R. H. and W. P. DUNLAP (1988) “Influence of chroma and hue on spatial balance of color pairs,” *Color Research & Application*, **13**(6), pp. 385–388.
- [73] NISHIYAMA, M., T. OKABE, I. SATO, and Y. SATO (2011) “Aesthetic quality classification of photographs based on color harmony,” in *Proceedings of the Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 33–40.
- [74] SAXENA, A., M. SUN, and A. Y. NG (2009) “Make3d: Learning 3d scene structure from a single still image,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **31**(5), pp. 824–840.

- [75] DENG, Y. and B. MANJUNATH (2001) “Unsupervised segmentation of color-texture regions in images and video,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23**(8), pp. 800–810.
- [76] LI, J. (2011) “Agglomerative connectivity constrained clustering for image segmentation,” *Statistical Analysis and Data Mining*, **4**(1), pp. 84–99.
- [77] BORJI, A. and L. ITTI (2013) “State-of-the-art in visual attention modeling,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **35**(1), pp. 185–207.
- [78] LIU, T., J. SUN, N. NING ZHENG, X. TANG, and H. YEUNG SHUM (2007) “Learning to Detect A Salient Object,” in *Proceedings of the Conference on Computer Vision and Pattern Recognition*, IEEE.
- [79] JUDD, T., K. EHINGER, F. DURAND, and A. TORRALBA (2009) “Learning to predict where humans look,” in *Proceedings of International Conference on Computer Vision*, IEEE, pp. 2106–2113.
- [80] ACHANTA, R., S. HEMAMI, F. ESTRADA, and S. SUSSTRUNK (2009) “Frequency-tuned salient region detection,” in *Proceedings of the Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 1597–1604.
- [81] JIANG, H., J. WANG, Z. YUAN, Y. WU, N. ZHENG, and S. LI (2013) “Salient object detection: A discriminative regional feature integration approach,” in *Proceedings of the Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 2083–2090.
- [82] HAREL, J., C. KOCH, and P. PERONA (2007) “Graph-based visual saliency,” *Advances in Neural Information Processing Systems*, **19**, pp. 545–552.
- [83] GOFERMAN, S., L. ZELNIK-MANOR, and A. TAL (2012) “Context-aware saliency detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **34**(10), pp. 1915–1926.
- [84] ERDEM, E. and A. ERDEM (2013) “Visual saliency estimation by nonlinearly integrating features using region covariances,” *Journal of Vision*, **13**(4), pp. 11–11.
- [85] DUAN, L., C. WU, J. MIAO, L. QING, and Y. FU (2011) “Visual saliency detection by spatially weighted dissimilarity,” in *Proceedings of the Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 473–480.
- [86] RUEOPAS, W., S. LEELHAPANTU, and T. H. CHALIDABHONGSE (2016) “A corner-based saliency model,” in *Proceedings of International Joint Conference on Computer Science and Software Engineering*, IEEE, pp. 1–6.



- [87] TAVAKOLI, H. R., E. RAHTU, and J. HEIKKILÄ (2011) “Fast and efficient saliency detection using sparse sampling and kernel density estimation,” in *Proceedings of Scandinavian Conference on Image Analysis*, Springer, pp. 666–675.
- [88] MURRAY, N., M. VANRELL, X. OTAZU, and C. A. PARRAGA (2011) “Saliency estimation using a non-parametric low-level vision model,” in *Proceedings of the Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 433–440.
- [89] HOU, X., J. HAREL, and C. KOCH (2012) “Image signature: Highlighting sparse salient regions,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **34**(1), pp. 194–201.
- [90] TAVAKOLI, H. R. and J. LAAKSONEN (2016) “Bottom-Up Fixation Prediction Using Unsupervised Hierarchical Models,” in *Proceedings of Asian Conference on Computer Vision*, Springer, pp. 287–302.
- [91] LI, G. and Y. YU (2015) “Visual saliency based on multiscale deep features,” in *Proceedings of the Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 5455–5463.
- [92] PAN, J., E. SAYROL, X. GIRO-I NIETO, K. MCGUINNESS, and N. E. O’CONNOR (2016) “Shallow and deep convolutional networks for saliency prediction,” in *Proceedings of the Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 598–606.
- [93] WUNDT, W. M. and C. H. JUDD (1907) *Outlines of Psychology*, W. Engelmann.
- [94] EKMAN, P. (1992) “An argument for basic emotions,” *Cognition & Emotion*, **6**(3-4), pp. 169–200.
- [95] EKMAN, P. and D. CORDARO (2011) “What is meant by calling emotions basic,” *Emotion Review*, **3**(4), pp. 364–370.
- [96] WATSON, D. and A. TELLEGEN (1985) “Toward a consensual structure of mood.” *Psychological Bulletin*, **98**(2), p. 219.
- [97] THAYER, R. E. (1989) *The Biopsychology of Mood and Arousal*, Oxford University Press.
- [98] ELLSWORTH, P. C. and K. R. SCHERER (2003) “Appraisal processes in emotion,” in *Handbook of Affective Sciences*, vol. 572, Oxford Press., pp. 572–595.

- [99] SCHERER, K., E. DAN, and A. FLYKT (2006) “What determines a feeling’s position in affective space? A case for appraisal,” *Cognition & Emotion*, **20**(1), pp. 92–113.
- [100] SILVIA, P. J. (2005) “Cognitive appraisals and interest in visual art: Exploring an appraisal theory of aesthetic emotions,” *Empirical Studies of the Arts*, **23**(2), pp. 119–133.
- [101] LEDER, H., B. BELKE, A. OEBERST, and D. AUGUSTIN (2004) “A model of aesthetic appreciation and aesthetic judgments,” *British Journal of Psychology*, **95**(4), pp. 489–508.
- [102] GRADY, C. and M. LEASE (2010) “Crowdsourcing document relevance assessment with Mechanical Turk,” in *Proceedings of the NAACL HLT Workshop on Creating Speech and Language Data with Amazon’s Mechanical Turk*, Association for Computational Linguistics, Stroudsburg, PA, USA, p. 172–179.
- [103] EICKHOFF, C., C. G. HARRIS, A. P. DE VRIES, and P. SRINIVASAN (2012) “Quality through flow and immersion: gamifying crowdsourced relevance assessments,” in *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, pp. 871–880.
- [104] BLANCO, R., H. HALPIN, D. M. HERZIG, P. MIKA, J. POUND, H. S. THOMPSON, and T. TRAN DUC (2011) “Repeatable and reliable search system evaluation using crowdsourcing,” in *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, pp. 923–932.
- [105] SARWAR, S. M., J. FOLEY, and J. ALLAN (2018) “Term Relevance Feedback for Contextual Named Entity Retrieval,” in *Proceedings of the Conference on Human Information Interaction & Retrieval*, ACM, pp. 301–304.
- [106] BENOIT, K., D. CONWAY, B. E. LAUDERDALE, M. LAVER, and S. MIKHAYLOV (2016) “Crowd-sourced text analysis: Reproducible and agile production of political data,” *American Political Science Review*, **110**(2), pp. 278–295.
- [107] MCDUFF, D. and M. SOLEYMANI (2017) “Large-scale Affective Content Analysis: Combining Media Content Features and Facial Reactions,” in *Proceedings of IEEE International Conference on Automatic Face & Gesture Recognition*, IEEE, pp. 339–345.

- [108] BUSO, C., S. PARTHASARATHY, A. BURMANIA, M. ABDELWAHAB, N. SADOUGHI, and E. M. PROVOST (2017) “MSP-IMPROV: An acted corpus of dyadic interactions to study emotion perception,” *IEEE Transactions on Affective Computing*, **8**(1), pp. 67–80.
- [109] ÇANO, E. and M. MORISIO (2017) “Crowdsourcing Emotions in Music Domain,” *International Journal of Artificial Intelligence & Applications*, **8**(4), pp. 25–40.
- [110] CHUA, T.-S., J. TANG, R. HONG, H. LI, Z. LUO, and Y. ZHENG (2009) “NUS-WIDE: a real-world web image database from National University of Singapore,” in *Proceedings of the International Conference on Image and Video Retrieval*, p. 48.
- [111] HUISKES, M. J. and M. S. LEW (2008) “The MIR Flickr retrieval evaluation,” in *Proceedings of The International Conference on Multimedia*, ACM, pp. 39–43.
- [112] REN, J., X. SHEN, Z. LIN, R. MECH, and D. J. FORAN (2017) “Personalized Image Aesthetics,” in *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pp. 638–647.
- [113] BUHRMESTER, M., T. KWANG, and S. D. GOSLING (2011) “Amazon’s Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data?” *Perspectives on Psychological Science*, **6**(1), pp. 3–5.
- [114] PAOLACCI, G., J. CHANDLER, and P. G. IPEIROTIS (2010) “Running Experiments on Amazon Mechanical Turk,” *Judgment and Decision Making*, **5**(5).
- [115] FORD, J. B. (2017) “Amazon’s Mechanical Turk: a comment,” *Journal of Advertising*, **46**(1), pp. 156–158.
- [116] GOODMAN, J. K., C. E. CRYDER, and A. CHEEMA (2013) “Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples,” *Journal of Behavioral Decision Making*, **26**(3), p. 213–224.  
URL <http://onlinelibrary.wiley.com/doi/10.1002/bdm.1753/abstract>
- [117] KEES, J., C. BERRY, S. BURTON, and K. SHEEHAN (2017) “An analysis of data quality: professional panels, student subject pools, and Amazon’s mechanical turk,” *Journal of Advertising*, **46**(1), pp. 141–155.
- [118] BRABHAM, D. C. (2008) “Crowdsourcing as a Model for Problem Solving An Introduction and Cases,” *Convergence: The International Journal of Research into New Media Technologies*, **14**(1), pp. 75–90.

- [119] SABATINELLI, D., T. FLAISCH, M. M. BRADLEY, J. R. FITZSIMMONS, and P. J. LANG (2004) “Affective picture perception: gender differences in visual cortex?” *Neuroreport*, **15**(7), pp. 1109–1112.
- [120] LITHARI, C., C. FRANTZIDIS, C. PAPADELIS, A. B. VIVAS, M. KLADOS, C. KOURTIDOU-PAPADELI, C. PAPPAS, A. IOANNIDES, and P. BAMIDIS (2010) “Are females more responsive to emotional stimuli? A neurophysiological study across arousal and valence dimensions,” *Brain Topography*, **23**(1), pp. 27–40.
- [121] COURSARIS, C. K., S. J. SWIERENGA, and E. WATRALL (2008) “An empirical investigation of color temperature and gender effects on web aesthetics,” *Journal of Usability Studies*, **3**(3), pp. 103–117.
- [122] BONNARDEL, N., A. PIOLAT, and L. LE BIGOT (2011) “The impact of colour on Website appeal and users’ cognitive processes,” *Displays*, **32**(2), pp. 69–80.
- [123] BOTTOMLEY, P. A. and J. R. DOYLE (2006) “The interactive effects of colors and products on perceptions of brand logo appropriateness,” *Marketing Theory*, **6**(1), pp. 63–83.
- [124] MCINNIS, J. H. and J. K. SHEARER (1964) “Relationship between color choice and selected preferences of the individual,” *Journal of Home Economics*, **56**(3), pp. 181–187.
- [125] WEINMANN, M., C. SCHNEIDER, and S. ROBBA-BISSANTZ (2013) “A Taxonomy of Web Personalization,” in *Proceedings of the Americas Conference on Information Systems*.
- [126] REINECKE, K. and A. BERNSTEIN (2011) “Improving performance, perceived usability, and aesthetics with culturally adaptive user interfaces,” *ACM Transactions on Computer-Human Interaction*, **18**(2), p. 8.
- [127] CYR, D., C. BONANNI, J. BOWES, and J. ILSEVER (2005) “Beyond trust: Web site design preferences across cultures,” *Journal of Global Information Management*, **13**(4), pp. 24–52.
- [128] KANG, K. and L. KOVACEVIC (2012) “The Effect of Culture on Emotions and Trust of Websites,” *Journal of Internet and e-Business Studies*, **2012**, pp. 1–12.
- [129] MARCUS, A. I., R. C. MOORE, and R. J. CYR (2001) “The role of microtubules in guard cell function,” *Plant Physiology*, **125**(1), pp. 387–395.

- [130] GALATIS, B. and P. APOSTOLAKOS (2004) “The role of the cytoskeleton in the morphogenesis and function of stomatal complexes,” *New Phytologist*, **161**(3), pp. 613–639.
- [131] QU, Y., P. SONG, Y. HU, X. JIN, Q. JIA, X. ZHANG, L. CHEN, and Q. ZHANG (2018) “Regulation of stomatal movement by cortical microtubule organization in response to darkness and ABA signaling in Arabidopsis,” *Plant Growth Regulation*, **84**(3), pp. 467–479.
- [132] YI, H., Y. RUI, B. KANDEMIR, J. WANG, C. ANDERSON, and V. PURI (2018) “Mechanical Effects of Cellulose, Xyloglucan, and Pectins on Stomatal Guard Cells of Arabidopsis thaliana.” *Frontiers in Plant Science*, **9**, pp. 1566–1566.
- [133] HARALICK, R. M., S. R. STERNBERG, and X. ZHUANG (1987) “Image analysis using mathematical morphology,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **9**(4), pp. 532–550.
- [134] KOVESI, P. D. (2013), “MATLAB and Octave Functions for Computer Vision and Image Processing,” Available from: <<http://www.peterkovesi.com/matlabfns/>>.
- [135] SARDER, P. and A. NEHORAI (2006) “Deconvolution methods for 3-D fluorescence microscopy images,” *IEEE Signal Processing Magazine*, **23**(3), pp. 32–45.
- [136] PANKAJAKSHAN, P., G. ENGLER, L. BLANC-FÚRAUD, and J. ZERUBIA (2013) “Deconvolution and Denoising for Confocal Microscopy,” in *Modeling in Computational Biology and Biomedicine*, Springer, pp. 117–163.
- [137] BOULANGER, J., C. KERVRANN, P. BOUTHEMY, P. ELBAU, J.-B. SIBARITA, and J. SALAMERO (2010) “Patch-based nonlocal functional for denoising fluorescence microscopy image sequences,” *IEEE Transactions on Medical Imaging*, **29**(2), pp. 442–454.
- [138] WANG, Y., Q. DAI, Q. CAI, P. GUO, and Z. LIU (2013) “Blind deconvolution subject to sparse representation for fluorescence microscopy,” *Optics Communications*, **286**, pp. 60–68.
- [139] KEUPER, M., T. SCHMIDT, M. TEMERINAC-OTT, J. PADEKEN, P. HEUN, O. RONNEBERGER, and T. BROX (2013) “Blind deconvolution of widefield fluorescence microscopic data by regularization of the optical transfer function (OTF),” in *Proceedings of the Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 2179–2186.

- [140] BAL, U., M. ENGIN, and U. UTZINGER (2013) “A multiresolution approach for enhancement and denoising of microscopy images,” *Signal, Image and Video Processing*, pp. 1–13.
- [141] MANNIESING, R., M. A. VIERGEVER, and W. J. NIESSEN (2006) “Vessel enhancing diffusion: A scale space representation of vessel structures,” *Medical Image Analysis*, **10**(6), pp. 815–825.
- [142] MENDRIK, A. M., E.-J. VONKEN, A. RUTTEN, M. A. VIERGEVER, and B. VAN GINNEKEN (2009) “Noise reduction in computed tomography scans using 3-D anisotropic hybrid diffusion with continuous switch,” *IEEE Transactions on Medical Imaging*, **28**(10), pp. 1585–1594.
- [143] PETROPOULOS, A. E., G. F. VLACHOPOULOS, S. G. SKIADOPOULOS, A. N. KARAHALIOU, and L. I. COSTARIDOU (2013) “Improving image quality in dual energy CT by edge-enhancing diffusion denoising,” in *Proceedings of the IEEE International Conference on Bioinformatics and Bioengineering*, IEEE, pp. 1–4.
- [144] ZHOU, X., C. YANG, and W. YU (2012) “Automatic mitral leaflet tracking in echocardiography by outlier detection in the low-rank representation,” in *Proceedings of the Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 972–979.
- [145] NADAKUDITI, R. R. (2014) “OptShrink: An algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage,” *IEEE Transaction on Information Theory*, **60**(5), pp. 3002–3018.
- [146] WEICKERT, J. (1997) “A Review of Nonlinear Diffusion Filtering,” in *Scale-Space Theory in Computer Vision*, vol. 1252, pp. 1–28.
- [147] PERONA, P. and J. MALIK (1990) “Scale-space and edge detection using anisotropic diffusion,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**, pp. 629–639.
- [148] ACAR, B. (2006) “Diffusion filters in image Processing: A brief review,” in *Similar NoE Tensor Workshop*, Las Palmas, Spain.
- [149] NIEMEIJER, M., J. STAAL, B. VAN GINNEKEN, M. LOOG, and M. D. ABRAMOFF (2004) “Comparative study of retinal vessel segmentation methods on a new publicly available database,” in *SPIE Medical Imaging 2004: Image Processing*, vol. 5370, International Society for Optics and Photonics, pp. 648–657.

- [150] BADSHA, S., A. W. REZA, K. G. TAN, and K. DIMYATI (2013) “A new blood vessel extraction technique using edge enhancement and object classification,” *Journal of Digital Imaging*, **26**(6), pp. 1107–1115.
- [151] CASELLES, V., R. KIMMEL, and G. SAPIRO (1997) “Geodesic Active Contours,” *International Journal of Computer Vision*, **22**(1), pp. 61–79.
- [152] WINK, O., W. NIESSEN, and M. VIERGEVER (2004) “Multiscale vessel tracking,” *IEEE Transactions on Medical Imaging*, **23**(1), pp. 130–133.
- [153] LI, H., T. SHEN, M. B. SMITH, I. FUJIWARA, D. VAVYLONIS, and X. HUANG (2009) “Automated actin filament segmentation, tracking and tip elongation measurements based on open active contour models,” in *Proceedings of the International Symposium on Biomedical Imaging: From Nano to Macro*, IEEE, pp. 1302–1305.
- [154] BÉRIAULT, S., Y. XIAO, D. L. COLLINS, and G. B. PIKE (2015) “Automatic SWI venography segmentation using conditional random fields,” *IEEE Transactions on Medical Imaging*, **34**(12), pp. 2478–2491.
- [155] LU, P., J. XIA, Z. LI, J. XIONG, J. YANG, S. ZHOU, L. WANG, M. CHEN, and C. WANG (2016) “A vessel segmentation method for multi-modality angiographic images based on multi-scale filtering and statistical models,” *Biomedical Engineering Online*, **15**(1), p. 120.
- [156] BATES, R., B. IRVING, B. MARKELC, J. KAEPLER, G. BROWN, R. J. MUSCHEL, M. BRADY, V. GRAU, and J. A. SCHNABEL (2019) “Segmentation of Vasculature from Fluorescently Labeled Endothelial Cells in Multi-Photon Microscopy Images,” *IEEE Transactions on Medical Imaging*, **38**(1), pp. 1–10.
- [157] MENG, Q., T. KITASAKA, Y. NIMURA, M. ODA, J. UENO, and K. MORI (2017) “Automatic segmentation of airway tree based on local intensity filter and machine learning technique in 3D chest CT volume,” *International Journal of Computer Assisted Radiology and Surgery*, **12**(2), pp. 245–261.
- [158] GU, L. and L. CHENG (2015) “Learning to boost filamentary structure segmentation,” in *Proceedings of International Conference on Computer Vision*, IEEE, pp. 639–647.
- [159] TANG, S., T. LIN, J. YANG, J. FAN, D. AI, and Y. WANG (2015) “Retinal vessel segmentation using supervised classification based on multi-scale vessel filtering and Gabor wavelet,” *Journal of Medical Imaging and Health Informatics*, **5**(7), pp. 1571–1574.

- [160] RICCI, E. and R. PERFETTI (2007) “Retinal blood vessel segmentation using line operators and support vector classification,” *IEEE Transactions on Medical Imaging*, **26**(10), pp. 1357–1365.
- [161] FU, H., Y. XU, S. LIN, D. W. K. WONG, and J. LIU (2016) “Deepvessel: Retinal vessel segmentation via deep learning and conditional random field,” in *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention*, Springer, pp. 132–139.
- [162] LI, Q., B. FENG, L. XIE, P. LIANG, H. ZHANG, and T. WANG (2016) “A cross-modality learning approach for vessel segmentation in retinal images,” *IEEE Transactions on Medical Imaging*, **35**(1), pp. 109–118.
- [163] MANINIS, K.-K., J. PONT-TUSET, P. ARBELÁEZ, and L. VAN GOOL (2016) “Deep retinal image understanding,” in *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention*, Springer, pp. 140–148.
- [164] WANG, S., Y. YIN, G. CAO, B. WEI, Y. ZHENG, and G. YANG (2015) “Hierarchical retinal blood vessel segmentation based on feature and ensemble learning,” *Neurocomputing*, **149**, pp. 708–717.
- [165] REDMON, J., S. DIVVALA, R. GIRSHICK, and A. FARHADI (2016) “You only look once: Unified, real-time object detection,” in *Proceedings of the Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 779–788.
- [166] RUSSAKOVSKY, O., J. DENG, H. SU, J. KRAUSE, S. SATHEESH, S. MA, Z. HUANG, A. KARPATY, A. KHOSLA, M. BERNSTEIN, ET AL. (2015) “Imagenet large scale visual recognition challenge,” *International Journal of Computer Vision*, **115**(3), pp. 211–252.
- [167] TIAN, Z., W. HUANG, T. HE, P. HE, and Y. QIAO (2016) “Detecting text in natural image with connectionist text proposal network,” in *Proceedings of the European Conference on Computer Vision*, Springer, pp. 56–72.
- [168] ZHOU, Z., F. FARHAT, and J. Z. WANG (2017) “Detecting dominant vanishing points In natural scenes with application to composition-sensitive image retrieval,” *IEEE Transactions on Multimedia*, **19**(12), pp. 2651–2665.
- [169] AVERILL, J. R. (1975) *A Semantic Atlas of Emotional Concepts*, APA.
- [170] YIK, M., J. A. RUSSELL, and J. H. STEIGER (2011) “A 12-point circumplex structure of core affect.” *Emotion*, **11**(4), pp. 705–731.



- [171] SEVA, R. R., K. G. T. GOSIACO, M. C. E. D. SANTOS, and D. M. L. PANGILINAN (2011) “Product design enhancement using apparent usability and affective quality,” *Applied Ergonomics*, **42**(3), pp. 511–517.
- [172] LU, X., P. SURYANARAYAN, R. B. ADAMS JR, J. LI, M. G. NEWMAN, and J. Z. WANG (2012) “On shape and the computability of emotions,” in *Proceedings of The International Conference on Multimedia*, ACM, pp. 229–238.
- [173] THÜRING, M. and S. MAHLKE (2007) “Usability, aesthetics and emotions in human–technology interaction,” *International Journal of Psychology*, **42**(4), pp. 253–264.
- [174] PEER, E., L. BRANDIMARTE, S. SAMAT, and A. ACQUISTI (2017) “Beyond the Turk: Alternative platforms for crowdsourcing behavioral research,” *Journal of Experimental Social Psychology*, **70**, pp. 153–163.
- [175] LU, X., R. B. ADAMS, J. LI, M. G. NEWMAN, and J. Z. WANG (2017) “An investigation into three visual characteristics of complex scenes that evoke human emotion,” in *Proceedings of the International Conference on Affective Computing and Intelligent Interaction*, IEEE, pp. 440–447.
- [176] SEGALIN, C., A. PERINA, M. CRISTANI, and A. VINCIARELLI (2017) “The pictures we like are our image: continuous mapping of favorite pictures into self-assessed and attributed personality traits,” *IEEE Transactions on Affective Computing*, **8**(2), pp. 268–285.
- [177] BLEI, D. M., A. Y. NG, and M. I. JORDAN (2003) “Latent dirichlet allocation,” *Journal of Machine Learning Research*, **3**, pp. 993–1022.
- [178] ISHIHARA, K., P. A. NGUYEN, A. C. GROEN, C. M. FIELD, and T. J. MITCHISON (2014) “Microtubule nucleation remote from centrosomes may explain how asters span large cells,” *Proceedings of the National Academy of Sciences*, **111**(50), pp. 17715–17722.
- [179] LECUN, Y., Y. BENGIO, and G. HINTON (2015) “Deep learning,” *Nature*, **521**(7553), p. 436.
- [180] LITJENS, G., T. KOOI, B. E. BEJNORDI, A. A. A. SETIO, F. CIOMPI, M. GHAFORIAN, J. A. VAN DER LAAK, B. VAN GINNEKEN, and C. I. SÁNCHEZ (2017) “A survey on deep learning in medical image analysis,” *Medical Image Analysis*, **42**, pp. 60–88.
- [181] RUEDEN, C. T., J. SCHINDELIN, M. C. HINER, B. E. DEZONIA, A. E. WALTER, E. T. ARENA, and K. W. ELICEIRI (2017) “ImageJ2: ImageJ for

- the next generation of scientific image data,” *BMC Bioinformatics*, **18**(1), p. 529.
- [182] ZHANG, Y., B. J. MATUSZEWSKI, L.-K. SHARK, and C. J. MOORE (2008) “Medical image segmentation using new hybrid level-set method,” in *Proceedings of The International Conference BioMedical Visualization: Information Visualization in Medical and Biomedical Informatics*, IEEE, pp. 71–76.
- [183] CHAN, T. F. and L. A. VESE (2001) “Active contours without edges,” *IEEE Transactions on Image Processing*, **10**(2), pp. 266–277.
- [184] SCHAFER, R. W. ET AL. (2011) “What is a Savitzky-Golay filter,” *IEEE Signal Processing Magazine*, **28**(4), pp. 111–117.
- [185] CIGNONI, P., M. CALLIERI, M. CORSINI, M. DELLEPIANE, F. GANOVELLI, and G. RANZUGLIA (2008) “Meshlab: an open-source mesh processing tool.” in *Eurographics Italian Chapter Conference*, vol. 2008, pp. 129–136.
- [186] TAUBIN, G. (1995) “A signal processing approach to fair surface design,” in *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques*, ACM, pp. 351–358.
- [187] LIU, J., J. M. WHITE, and R. M. SUMMERS (2010) “Automated detection of blob structures by hessian analysis and object scale,” in *Proceedings of the International Conference on Image Processing*, IEEE, pp. 841–844.
- [188] RUI, Y., C. XIAO, H. YI, B. KANDEMIR, J. Z. WANG, V. M. PURI, and C. T. ANDERSON (2017) “POLYGALACTURONASE INVOLVED IN EXPANSION3 functions in seedling development, rosette growth, and stomatal dynamics in *Arabidopsis thaliana*,” *The Plant Cell*, **29**(10), pp. 2413–2432.
- [189] KASS, M., A. WITKIN, and D. TERZOPOULOS (1988) “Snakes: Active contour models,” *International Journal of Computer Vision*, **1**(4), pp. 321–331.
- [190] MOTULSKY, H. J. and R. E. BROWN (2006) “Detecting outliers when fitting data with nonlinear regression—a new method based on robust nonlinear regression and the false discovery rate,” *BMC Bioinformatics*, **7**(1), p. 123.

## **Vita**

### **Bariş Kandemir**

Bariş Kandemir was born in İzmir, Turkey on January 25, 1990. He received his B.Sc. degree with Dean's High Honor List in Electrical and Electronics Engineering from Boğaziçi University, İstanbul, Turkey in 2012. Bariş received his Ph.D. degree in Spring 2019 under supervision of Prof. James Z. Wang. His research interests lie in computational aesthetics and affect analysis, and biomedical image processing. During his Ph.D. study, he was a research exchange student at Key Laboratory of Intelligent Information Processing at Fudan University, Shanghai, China between May and July 2016. He was assigned as a graduate assistant for Institute for CyberSciences at Penn State during Spring 2017. He worked as a software intern for landmark automation team in DeepMap Inc. during Summer 2017. He was appointed as graduate teaching fellow for Fall 2017. In Summer 2018, he joined landmark automation team in DeepMap Inc. as a full time employee.