Computer Vision Applications of Head Motion and Cell Analysis

A THESIS

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL

OF THE UNIVERSITY OF MINNESOTA

BY

Jordan K. Hashemi

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

Master of Science

Adivsor: Prof. Tryphon T. Georgiou

Co-Advisor: Prof. Guillermo R. Sapiro

May, 2013

Acknowledgements

I owe my deepest gratitude to Professor Guillermo for advising me through my Master's study as well as be a driving force for me to want to purse this graduate degree. His guidance, support, and encouragement have not only made me a better researcher, but also a better person. I am truly grateful to have such an exceptional mentor and friend.

I would like express my gratitude to my committee members Profs. Tryphon Georgiou, Nikolaos Papanikolopoulos, and Mos Kaveh for their thoughtful suggestions and time. I would especially like to thank Prof. Georgiou for 'taking me under his wing' as a student during the last semester of my studies, after Prof. Sapiro moved out from the University of Minnesota. I would also like to thank the grants that financially support our research.

Throughout my studies, I have been fortunate enough to collaborate with a wide range of researchers from different fields: Amy Esler, Judith Berman, Ben Harrison, Vassilios Morellas, and Nikolaos Papanikolopoulos. They have provided me with insightful knowledge and ideas, and I have thoroughly enjoyed their company. I am especially grateful for the immeasurable effort Ben Harrison has spent fine-tuning part of my work.

The material in this thesis has greatly benefited from discussions and interactions with colleagues: Pablo Cancela, Alexey Castrodad, Marcelo Fioro, Jinyoung Kim, Thiago Spina, Pablo Sprechman, Zhongwei Tang, and Mariano Tepper. Thanks to them for all of their support and friendships. I would like to give a very special thanks to Mariano Tepper, for the immense amount of help and guidance he has provided me throughout

my studies. I cannot emphasize enough how thankful I am of him.

This work would not be possible if it was not for the everlasting support of my family: my mom JoEll, dad Zack, sister Dayna, and brother Connor. Finally, I would like to thank my girlfriend Jessica, for her patience and support.

Abstract

Computer vision approaches have been applied to a wide variety of domains. These approaches are able to replicate human processes as well as provide new insights. Once developed, computer vision approaches also permit high throughput analysis of processes that would otherwise be tedious. In this thesis, we focus on developing tools that replicate and go above human analysis for two applications. We aim at developing tools to track infants‘ head motions in a non-intrusive manner during an autism assessment. We propose a tracking based method to facial features from a single camera and use their changing locations to sequentially update the motions. We also aim at developing an elliptical-cell and multiple-fluorescence segmentation program to allow for the high throughput analysis of C.albicans and S.cerevisiae yeast cultures. We propose to segment pseudohyphal cells based on an extension of the circular Hough transform; as well as, segment multiple types of fluorescence labeling.

Contents

List of Tables

List of Figures

1 Introduction

Computer vision has been applied to a wide variety of domains ranging from tracking athletes in videos ([47] and references therein) to facial recognition in images ([22] and references therein). These examples are part of two in-depth areas of computer vision: motion analysis and object recognition. Motion analysis from a video can be defined as determining what is happening in a sequence of images; whereas object recognition can be viewed as determining what is happening in a single image.

Both definitions aim at answering the question 'what is happening?'; a question that is dependent on the specific application's need. Depending on the application and problem, sometimes computer vision approaches are developed to replicate how humans process information (i.e. track a person running), while sometimes they are developed to analyze information that humans cannot process (i.e. find a latent manifold that an individual's face is contained in). The beauty of computer vision approaches is that a majority of the time they are able to do both, replicate human processes as well as provide new insights. Once developed, computer vision approaches also permit high throughput analysis of processes that would otherwise be tedious. This in turn, allows for new information to be discovered through data mining. In this thesis, we focus on developing tools that replicate and go above human analysis for an application related to motion analysis: head motion estimation; and an application related to object recognition: yeast cell analysis.

1.1 Motion analysis: head motion estimation

Determining the motion of a person's head is learned at a very young age. Even children, who are trying to sneak a piece of cake, are able to understand what their parents are implying when their parents move their heads left then right, then left then right. However, developing computer vision tools to robustly estimate head motion remains a major challenge. Head motion and pose estimation involve three motions of the head: yaw (left and right), pitch (up and down), and roll (tilting movement). Recently, literature has been primarily focused on head pose estimation. Vatahska et. al. [44] developed a feature based head pose estimation where they detect specific facial features (the eyes, nose, mouth, and ears) in images, and trained neural networks to estimate the yaw, pitch, and roll angles. Their work achieved high accuracy for frontal faces but showed significantly worse performance for profile faces. Ferrari et. al. [32] published work where they estimated head pose in $3D$ from single images using Gaussian process regressions. Both of these works dealt with images, thus there were no constraints for temporal continuity as well as smooth motion constraints.

To ensure the smooth motion and continuity constraints for head motion estimation, we consider implementing a tracking method similar to the broad frameworks provided in [35]. Other tracking methods, such as those based off of [1, 31, 46], create, train, and fit 3D head models. These methods show very promising results for frontal faces; however, performance varies during instances where the face transitions between frontal and profile. This is partially due to the fact that when the face is in a frontal view, it exhibits symmetrical features (e.g. the eyes are roughly equal distances from the nose). As the face transitions to a profile view, this face symmetry aspect is lost. Thus to accommodate for these face-appearance transitions, we propose a head motion tracking approach that focuses on only part of the face. In Chapter 2, we propose and present results of our approach for estimating infants' head motions during specific autism assessment tasks

and activities.

1.2 Object recognition: cell and fluorescence segmentation

An image containing a colony of densely packed yeast cells can be a frightening sight for a science undergraduate student who has to hand-label each cell and determine their shape characteristics. Luckily, there have been major advancements in cell segmentation approaches over the years [2, 4, 14, 27, 29, 48]; however most approaches compromise between accuracy and robustness. Cell segmentation methods are highly dependent on the imaging setup and environment. For the most part, the methods and approaches are developed based on a specific labs‘ needs and because of this, new biological protocols are being developed that are based on applied computer vision approaches. Nearly twenty years ago, Prasher et. al. [39] successfully cloned the green fluorescent protein. This work revolutionized the cell biology community as well as jump started the computer vision community to assist with biological analysis. Now, the GFP gene has been introduced in yeast strains allowing for a wide variety of analysis including the discovery of drug treatments [21]. Without computer vision approaches for biological applications, scientists would be restricted to either meticulously hand-label cells in images or be forced to alter their protocols to allow for a more reasonable analysis.

One of the aims for this thesis is based on developing an accurate, automatic, robust, and user-friendly cell analysis program that combines both cell and fluorescence segmentations of two widely used yeast strains. In Chapter 3, we present our developed program and demonstrate some of its multiple applications.

1.3 Document organization

One of the particular goals of this thesis is to develop tools for immediate real-world applications. The works within this thesis, were done in collaboration with University of Minnesota's Computer Science, Pediatrics, Psychology, Microbiology Departments. Not only do we present motivations for two different applications belonging to computer vision, we also validate our developed tools and demonstrate their value across many research communities including the psychology and cell biology communities.

The rest of the thesis is organized as follows. In Chapter 2, we present tools for estimating head motion to aid in the assessment of specific autism assessment tasks and activities [20, 18, 19]. Then in Chapter 3, we present a program for high throughput yeast cell analysis [17, 16, 15]. For both Chapter 2 and 3, we first present the motivation for the tools, then we propose the developed tools, and lastly we show results of our tools for the specific application. Finally, conclusions are drawn in Chapter 4.

2 Head motion estimation of infants during autism assessment

2.1 Chapter summary

In this chapter we develop computer vision tools to assist the assessment of the Autism Observation Scale for Infants (AOSI). In particular, we develop a non-intrusive head motion estimation method to measure critical AOSI tasks and activities that assess visual attention. Our method consists of tracking specific facial features to estimate yaw (left and right) and pitch (up and down) head motions. We present results from three AOSI tasks and activities and compare our results with expert and non-expert clinicians, demonstrating that the proposed system provides insightful knowledge to augment the clinician's behavioral observations obtained from real in-clinic assessments. The work in this chapter is cited from [20, 18, 19].

2.2 Introduction and motivation

The analysis of children's natural behavior is of key importance for the early detection of developmental disorders such as autism spectrum disorder (ASD). For example, several studies have revealed behaviors indicative of ASD in early home videos of children that were later diagnosed with ASD [49] and references therein. These studies involved video-recording these environments and then analyzing the data a posteriori,

using frame-by-frame viewing by an observer who typically trains for several weeks to achieve inter-rater reliability. Hours of labor are required, thereby making such analyses impractical for clinical settings as well as for big data studies aiming at the discovery or improvement of behavioral markers. While clinical tools for early diagnosis of ASD are available, they require administration and interpretation by specialists. Most families lack easy access to specialists in ASD; for example, the waiting list for an evaluation at the leading ASD Clinic at the University of Minnesota is 6 months for children age 4 and under. There is a need for automatic and quantitative analysis tools that can be used by general practitioners in child development, and in general environments, to identify children at-risk for ASD and other developmental disorders.

Although much is unknown about the underlying causes of ASD, it is characterized by abnormalities in social interactions and communication and the presence of restricted repetitive behaviors [49]. Neuropathological studies indicate that ASD has its origins in abnormal brain development early in prenatal life [40]. Moreover, [49] argue that many children with ASD exhibit several specific behavioral markers as early as in the first year of life. These markers appear, among others, in activities involving visual attention, often expressed as difficulties in disengagement and shifting of attention [30]. Such markers motivate the system developed and reported in this paper.

Despite this evidence, the average age of ASD diagnosis in the US is 5 years [42]. Recently, much research and clinical trials have focused on early diagnosis to allow for early intensive intervention. Early intervention, initiated in preschool and sustained for at least 2 years, can substantially improve child outcomes [7]. Detecting ASD risk and starting interventions before the full set of behavioral symptoms appears has an even greater impact, preventing difficult behaviors and delayed developmental trajectories from taking hold [7]. Early diagnosis is achieved by following a comprehensive battery of developmental and behavioral tests and parent interviews, with the goal of detecting

behavioral symptoms consistent with ASD. However, few specialized clinics exist to offer these assessments to the very young. Improving availability of early diagnosis may be achieved by developing screening tools that can be used in regular pediatric clinics and school environments, thereby reaching a large population very early. Towards this end, we develop semi-automatic computer vision video analysis techniques to aid in early detection.

These tools aid the practitioner in the diagnosis task by providing accurate and objective measurements. These measurements can further provide means for improving the shareability of clinical records without compromising anonymity. In addition, and particularly for research, automatic analysis will permit researchers to analyze vast amounts of naturally recorded videos, opening the door for data mining towards the improvement of current assessment protocols and the discovery of new behavioral features. As opposed to other research projects [23, 12, 28], where artificial setups are used, one of our main goals is to provide non-intrusive capturing systems that do not necessarily induce behavioral modification in the children. In other words, hardware must not constrain the testing environment: the clinician is free to adjust testing conditions as needed, and children are not asked to wear any type of sensors [13, 36] or perform any non-natural tasks.

The results in this work are from actual clinical recordings, in which the at-risk infant/toddler is tested by an experienced clinician following the Autism Observation Scale for Infants (AOSI) [3]. The AOSI is a well-validated behavioral observation tool [9] for gathering information on early ASD risk signs. It involves a set of semi-structured activities that provide an interactive context in which the examiner engages the infant in play while conducting a set of systematic presses to elicit specific child behaviors. Through the development of the AOSI, [49] identified multiple behavioral markers for early detection of ASD. The AOSI is divided into two main parts/categories: (1) a set of tabulated

tasks that are designed for assessing specific behaviors, where each task consists of a certain number of presses and the child's responses receive scores; (2) a freeplay session, in which the clinician assesses the social behavior of the child while he/she is allowed to explore toys/objects. Throughout the freeplay session, the clinician extracts many behavioral measurements. The studied behaviors mainly include social interactions and thus their assessments also take place throughout the entire session. The less structured nature of these holistic assessments makes the development of automated tools more challenging.

In this work, we focus on computer vision tools for two AOSI behaviors belonging to the first category, namely Disengagement of Attention and Visual Tracking. According to the AOSI, Disengagement of Attention is characterized as the "ability to disengage and move eyes/attention from one of two competing visual stimuli" [3], while Visual Tracking is characterized as the "ability to visually follow a moving object laterally across the midline" [3]. We also demonstrate, that the computer vision tools here developed can be of use in a more complex scenario, such as the freeplay session. In particular, we explore one activity belonging to the freeplay session, the ball playing activity. In the AOSI, the shared interest behavioral marker is characterized as the "ability to use eyes to reference and share interest in an object or event with another person" ([3], also known as "Social interest and Shared affect"), and is partially assessed during the ball playing activity.

The proposed method for head motion estimation consists of tracking specific facial features and using their changing locations to estimate yaw (left and right) and pitch (up and down) head motions. In Section 2.3 we provide detailed explanations of both the facial feature tracking and head motion estimation techniques. We examine the results of our method for the three stated AOSI tasks and activities as well as compare our results with that of ASD expert and non-expert clinicians. Please note the displayed

images in this chapter are downsampled, blurred, and/or partially blocked to preserve anonymity (processing was done on the original videos). Approval for this work was obtained from the Institutional Review Board at the University of Minnesota.

2.3 Proposed head motion estimation method

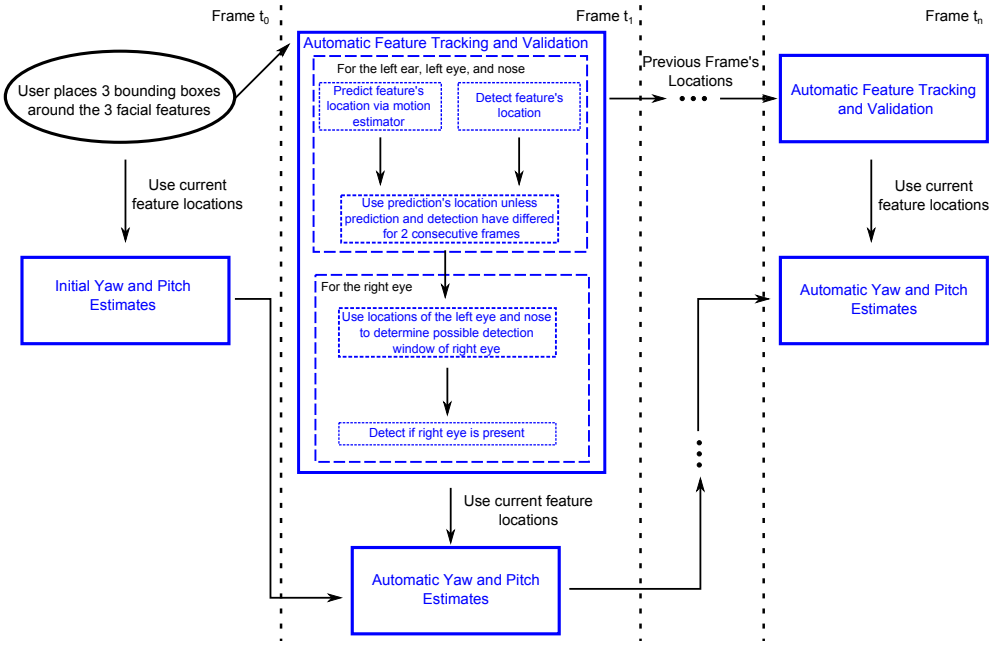


Figure 2.1: Data flow of the system for estimating yaw and pitch motions. We only require the user to place a bounding box around the left ear, left eye, and nose in the first frame of the video segment (black ellipse at time t_0). All the subsequent steps occur in a fully automatic fashion (blue blocks).

The proposed framework is consists of tracking three facial features: the nose, an ear and eye that correspond to the same side of the face. Please note that for simplicity, we will refer to the tracked ear and eye as the left ear and left eye. In instances where only the right side of the face is shown, we can easily rotate and flip the image so that it seems only the left side of the face is visible. We assume that, in the first frame, we have bounding boxes of the three facial features. These bounding boxes are in practice

selected by hand on the first frame. Estimations of yaw and pitch motions are calculated from the tracked positions of the facial features. Figure 2.1 provides an overview of our entire method.

2.3.1 Robust facial feature tracking

We aim at tracking the three facial features following a scheme loosely based on the TLD tracker [25]. We use dense motion estimation coupled with a validation step that employs an offline-trained facial feature detector. The dense motion estimator [43] tracks the features with high accuracy in most cases, but when the child's head moves quickly, illumination changes can sometimes cause the tracker to lag behind the features. Thus we validate the output of the tracker using facial feature detectors in every frame.

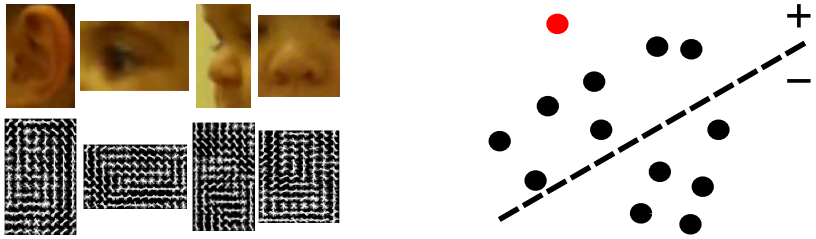


Figure 2.2: **Left images:**Examples of HOG descriptors for the left ear, left eye, nose in profile view, and nose in frontal view. The top row represents example image patches, while the bottom row shows the computed HOG descriptor with 9 orientation bins corresponding to the image above it. **Right plot:** Example of maximal output from a classifier. Circles above the dotted hyperplane represent positive detections while the circles below represent negative ones. The red circle corresponds to the detection with the maximal output from the classifier.

To validate the features we train left eye, right eye, left ear, and nose detectors. For this, we adapt the method by [6] see also [11, 44], proposed for pedestrian detection, to our particular setting. Our method employs multiscale Histograms of Orientated Gradients (HOG) using 8 x 8 pixel blocks and 9 orientation bins as descriptors to represent each facial feature (see left images in Figure 2.2), and then classifies these descriptors

using a Support Vector Machine[1] with a radial basis function kernel. The employed classifier determines the location of the facial feature based on the detection which exhibits a maximal response from the classifier (that is, the detected location is the extrema of each feature's classifier output see right plot in Figure 2.2). When viewing from a single, stationary camera, some facial features differ greatly in appearance between frontal and profile head views, e.g. in a profile view only one nostril is present, while in a frontal view both nostrils are present. To accommodate for these changes, we train two descriptors for detecting the nose (see left images in Figure 2.2). As positive training samples for the descriptors, we use hand labeled facial patches from children in our experimental environment. As negative training samples, we extract random patches from around multiple children's faces.

For each frame, search areas for the facial feature detectors are defined around the bounding boxes given by the tracker. The left eye, left ear, and nose are present in every frame for the given camera position and their final detected positions are determined by the locations that exhibit a maximal response from the classifier. The tracker's bounding boxes are validated if their centers are within the bounding boxes returned by the detectors; however, if the tracker's centers are outside of the detector's bounding boxes for two consecutive frames, then the corresponding bounding box for the tracker is reset to a new location within the detector's bounding box. Determining the presence of the right eye aids in the estimation of the yaw motion. The rectangular search area for the right eye, which is not tracked since it appears and disappears constantly due to the camera position, is based on the location of the detected nose and the horizontal and vertical distances between the detected left eye and nose. More specifically, the search area is between the detected nose's location plus/minus the horizontal and vertical distances between the detected left eye and nose. Also employed as a right eye search area

[1] `http://www.csie.ntu.edu.tw/~cjlin/libsvm/`

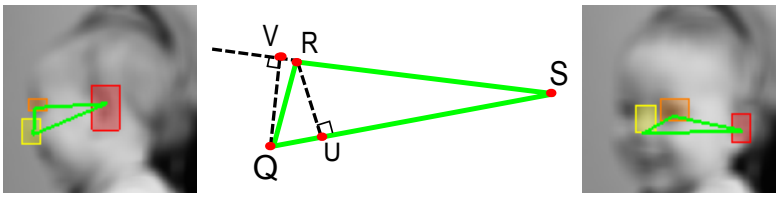


Figure 2.3: The triangle created by the left ear, left eye, and nose. The leftmost and rightmost images depict the triangle when the infant is looking right and more towards the camera, respectively. The middle image shows the points used for calculating $\widehat{\text{yaw}}$.

restriction is that the nose must be between the left and right eyes.

Thus, using our method we are able to track the facial features via a dense motion estimator, and validate their positions via the facial feature detectors. To estimate head motion, see below, we use the facial feature locations given by the dense motion estimators. The dense motion estimator provides smoother and more precise locations than the detector. The overall procedure is then both accurate and reliable.

2.3.2 Head motion estimation

In our setup, which is dictated by the clinician's preference, the child's face is predominantly in a profile view for the Sharing Interest activity. As a way to provide an accurate motion estimation of the pitch angle we cumulatively sum the vertical coordinate changes of the left eye and nose with respect to the left ear every frame. We expect a positive sum when the child is looking up and a negative sum when the child is looking down, the magnitude representing how much the child is looking up or down.

For estimating the yaw angle motion in the Visual Tracking and Disengagement of Attention activities, we calculate two ratios based on the triangle created by the left ear, left eye, and nose (Figure 2.3); we also use information about the presence of the right eye. Let Q, R, and S denote the locations of the nose, left eye, and left ear, respectively. For the first ratio $r_{\text{NoseToEye}}$, we project R into the line defined by QS, thus defining the point U; we then define $r_{\text{NoseToEye}} = |US|/|QS|$, where $|\cdot|$ is the Euclidian distance. For

the second ratio we project Q into the line defined by RS, defining $r_{\text{EyeToEar}} = |VR|/|RS|$.

The two ratios r_{EyeToEar} and $r_{\text{NoseToEye}}$ are inversely proportional. Looking at Figure 2.3 we can observe that when the face is looking in profile view, r_{EyeToEar} will be large and $r_{\text{NoseToEye}}$ will be small; conversely when the face is in frontal view (looking more towards the camera). To combine these two ratios into one value, we calculate the normalized difference between them, $\widehat{\text{yaw}} = \frac{r_{\text{EyeToEar}} - r_{\text{NoseToEye}}}{r_{\text{EyeToEar}} + r_{\text{NoseToEye}}}$. Thus, as the child is looking to his/her left, $\widehat{\text{yaw}}$ goes to -1; and as the child is looking to his/her right, $\widehat{\text{yaw}}$ goes to 1. The presence of the right eye further verifies that the infant is looking left.

We incorporate if the right eye is present or not to verify that the infant is looking left or right at the maximum and minimum $\widehat{\text{yaw}}$ values.

2.4 Experimental Results

Specific AOSI procedures

During the AOSI assessment, the clinician performs three trials for the Disengagement of Attention task and two trials for the Visual Tracking task, per participant. Every trial receives an AOSI-tabulated score, according to the following guidelines:

- Disengagement of Attention. This activity consists of (1) shaking a noisy toy to one side of the infant until his/her attention is engaged, and (2) then shaking a second noisy toy on the opposite side, while continuing to shake the first object. A delayed response is an ASD risk sign [30]. A trial is considered “passed” if the child looks to the second object in less than 1s, considered “delayed” if the child looks after a 1-2s delay, and considered “stuck” if the child looks after more than 2s.

- Visual Tracking. To evaluate this activity, the following is performed: (1) a rattle or other noisy toy is used to engage the infant’s attention, (2) the rattle is positioned

to one side of the infant, and (3) the rattle is then moved silently at eye level across the midline to the other side. Infants with ASD usually exhibit discontinuous and/or a noticeably delayed tracking [3]. Depending on how continuously and smoothly the participant is able to track the object, the trial is considered "passed," "delayed or interrupted," or "partial or no tracking."

The clinician makes a "live" judgment about these time frames or may look at videos of this task if available. Finally, an overall score for each task is computed by merging the individual trials. We followed the protocol of comparing the assessments done by: (1) an expert psychologist in autistic children, (2) a child/adolescent psychiatrist, (3) two psychology students with no particular autism training, and (4) the results of our new computational tools. The child/adolescent psychiatrist and two psychology students assigned their scores by following the AOSI guidelines, without prior training, while watching the videos used by the automatic method. This setup allows us to contrast the automatic method's findings with human assessments across the full range of expertise.

In the ball playing activity, a ball is rolled on the table towards the infant after engaging his/her attention. After receiving the ball, the clinician analyzes the child's ability to acknowledge the involvement of another person in the game-play by looking to either the clinician or the caregiver. Infrequent or limited looking to faces is an early ASD risk sign [49, 3].

In our clinical setup, we use a low-cost GoPro Hero HD color camera (with a resolution of 1080p at 30 fps), placed by the clinician on her table.

Participants

To validate our method, we applied it to a study that involves 12 participants, including both males and females ranging in age from 5 to 18 months. We have gathered our data from a series of ASD evaluation sessions of an ongoing concurrent study performed on a

group of at-risk infants, at the Department of Pediatrics of the University of Minnesota.

All participants were classified as a baby sibling of someone with ASD, a premature infant, or as a participant showing developmental delays. Table 2.1 presents a summary of this information.

Table 2.1: Information on Participants involved in this study. Each participant was chosen for a different reason: being a baby sibling of someone with ASD, a premature infant, or showing developmental delays.

Part #	Age (months)	Gender	Risk Degree
#1	14	F	Showing delays
#2	5	F	Baby sibling
#3	16	M	Showing delays
#4	15	M	Showing delays
#5	8	M	Premature infant
#6	9	F	Premature infant
#7	10	F	Premature infant
#8	9	M	Premature infant
#9	7	M	Premature infant
#10	6	M	Baby sibling
#11	9	M	Premature infant
#12	18	M	Showing delays

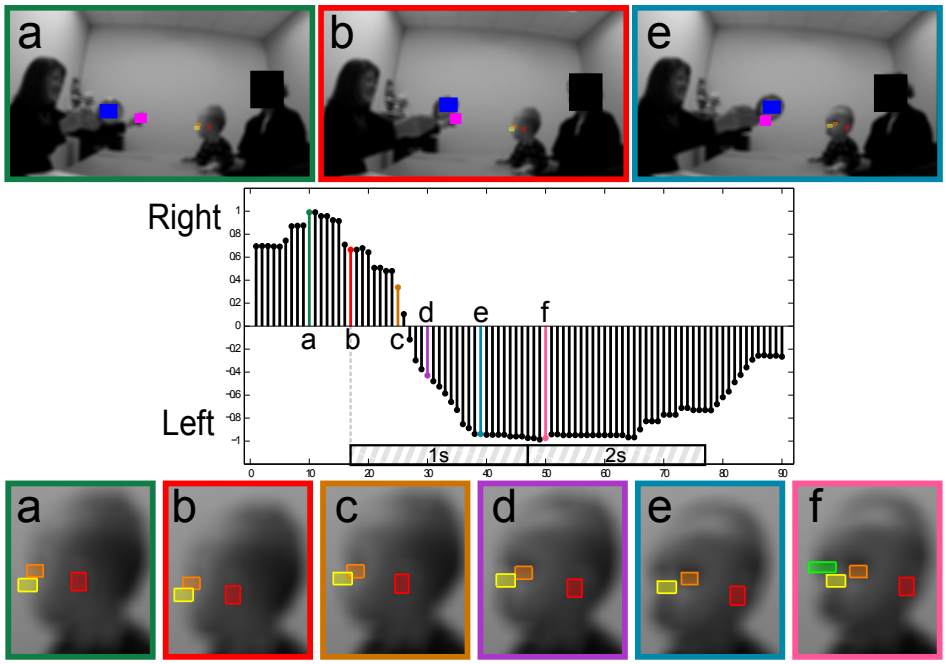
We acquired our data in this fashion as an initial effort towards developing our computer vision methods. Since we are interested in initially measuring known behavioral markers, the lack of a control group composed of children with typical development is mitigated.

2.4.1 Disengagement of attention

Table 2.2 summarizes the results of our method, the clinical assessments, and the ratings by a child/adolescent psychiatrist and two psychology students for the Disengagement of Attention task. After marking when the second object is presented, our method is able to automatically determine the delay from when the participant disengages from the

Table 2.2: Results of Disengagement of Attention task. Scores from the clinician (Clin.), automatic method (Automatic), a psychiatrist (Psy.), and two students (St. 1 and St. 2) for each trial. A trial is considered either as "passed" (Pass), "delayed" (Del), or "stuck" (Stck) depending on whether the child disengages from the first object in less than $1s$, between $1-2s$, or more than $2s$, respectively. We also present the automatically computed delay that the child takes to disengage. Note that we consider a $+\frac{1}{3}$ of a second margin for each delay to accommodate human error of making a live judgment (colors are added to facilitate the comparison).

Part.	Clin.	Automatic		Psy.	St. 1	St. 2
		Score	Delay (s)			
	First Trial Score					
#1	Pass	Pass	0.9	Pass	Del	Del
#2	N/A	Del	1.87	N/A	N/A	N/A
#3	Pass	Pass	0.5	Pass	Del	Del
#4	Pass	Pass	0.23	Pass	Pass	Pass
#5	Del	Pass	1.07	Del	Del	Del
#6	Pass	Pass	1.03	Pass	Del	Pass
#7	Pass	Pass	0.83	Pass	Del	Del
#8	-	-	-	-	-	-
#9	-	-	-	-	-	-
#10	Pass	Pass	0.87	Del	Del	Del
#11	Pass	Pass	0.83	Pass	Del	Pass
#12	Pass	Pass	0.93	Pass	Pass	Pass
	Second Trial Score					
#1	Pass	Pass	0.7	Pass	Pass	Pass
#2	-	-	-	-	-	-
#3	-	-	-	-	-	-
#4	Pass	Pass	1.1	Pass	Pass	Pass
#5	Del	Del	1.77	Del	Stck	Del
#6	Pass	Del	1.43	Pass	Pass	Pass
#7	Pass	Pass	0.97	Del	Del	Del
#8	Pass	Pass	1.33	Pass	Del	Del
#9	-	-	-	-	-	-
#10	Pass	Pass	1.3	Pass	Pass	Pass
#11	Pass	Pass	0.63	Pass	Pass	Pass
#12	Pass	Pass	0.9	Pass	Del	Del
	Third Trial Score					
#1	Pass	Pass	0.37	Pass	Pass	Pass
#2	-	-	-	-	-	-
#3	-	-	-	-	-	-
#4	Pass	Pass	0.3	Pass	Pass	Pass
#5	Pass	Pass	0.5	Pass	Del	Pass
#6	Pass	Pass	0.7	Del	Del	Del
#7	Pass	Pass	1.13	Pass	Del	Del
#8	-	-	-	-	-	-
#9	Del	Del	1.37	Pass	Del	Stck
#10	Pass	Pass	1.33	Del	Del	Del
#11	Pass	Pass	0.87	Pass	Pass	Pass
#12	Pass	Pass	0.87	Pass	Pass	Pass

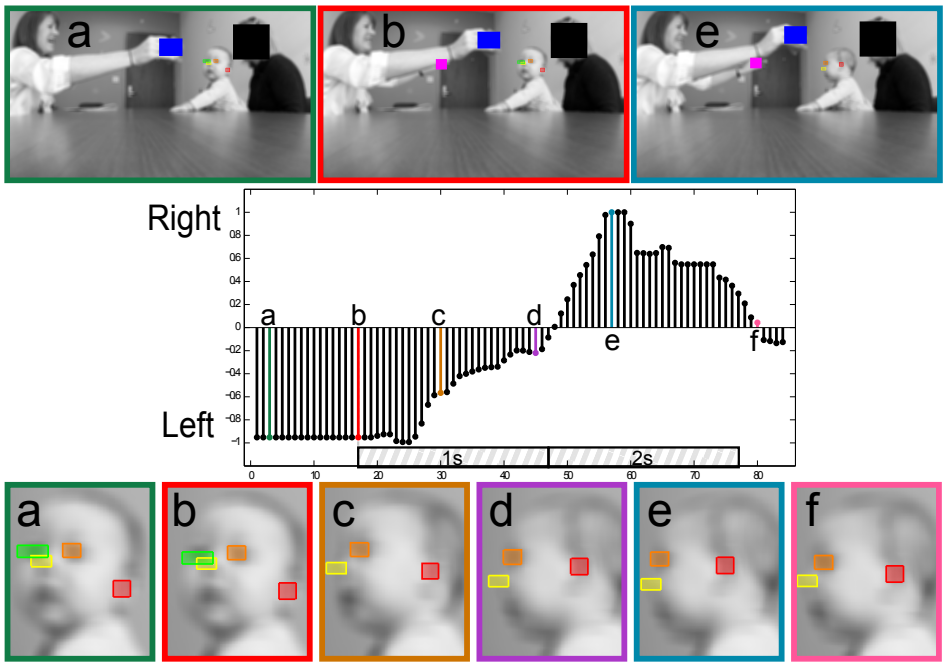


(a)

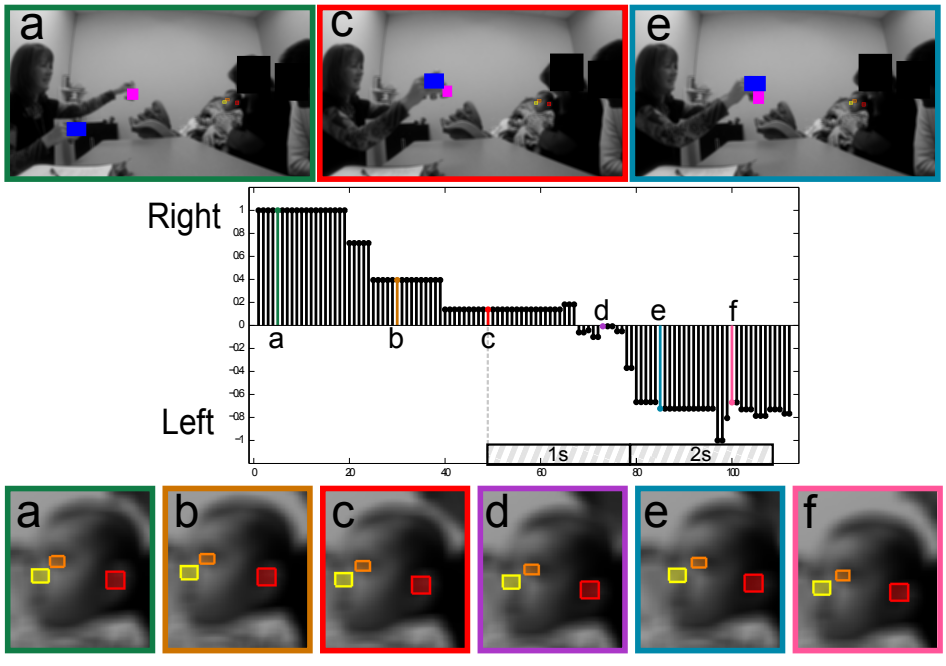
Figure 2.4: First example of the Disengagement of Attention task. **Top:** clinician holding one object, when the second object appears, and when the child recognizes the second object. **Middle:** changes in the yaw motion ($\widehat{\text{yaw}}$ values in the y-axis) for every frame (x-axis). The dotted line represents when the second object is presented, followed by boxes representing 1 and 2 seconds after the object is presented. **Bottom:** 6 examples of the infant's face during the task. All facial features are automatically detected and tracked (as indicated by the colored boxes around the nose, eyes and ear). Colors and letters identify corresponding images and spikes in the graph.

first object to the second. We present this delay in terms of how many seconds/frames it takes for the participant to disengage (note that we are recording the video at 30 frames per second). We incorporate a $+\frac{1}{3}$ of a second margin for each delay to accommodate human error of making a live judgment.

Since the current set up for the visual attention tasks only involves a single camera placed non-intrusively, there are trials that our current method cannot handle (this could be easily solved in the future with a second low-cost camera). These trials include instances when the participant left the camera's field of view or when a toy or object



(b)



(c)

Figure 2.4: (Continued) Two examples of the Disengagement of Attention task.

obstructed it. For Table 2.2, the trials with blank spaces and a horizontal line correspond to such cases. Out of the 24 trials that the clinician assigned a “pass” score, our method agreed on 23 of them and scored a “delayed” for the other trial. And out of the 3 trials the clinican scored “delayed” our method agreed on 2 trials, scoring one as a “pass.” The clinician did not assign a score for #2, stating that it was a “spoiled” trial due to the participant being afraid of the toys. However, we show our method’s results to exemplify a possible Disangagement of Attention score. Note that with the automatic method here proposed, we can achieve much higher accuracy then the one obtained in real-time clinicial assessments.

To further clarify our results, Figure 2.4 displays examples of our method’s results and important cases for the Disengagement of Attention task. In example 2.4(a), the participant is able to disengage from the first object and look at the second within 0.7s (21 frames) of the second object being presented. This would be scored as “passed” on the AOSI test. The participant in example 2.4(b) disengages to the second object within 1.3s (40 frames), which would be scored as “delayed” on the AOSI test. Not only does it take the third participant (example 2.4(c)) over 1s to look at the second object (which is “delayed” on the AOSI), but the participant displays piece-wise constant lateral head movements compared to the other two examples (which presented a much smoother motion), a pattern virtually impossible to detect with the naked eye. Again, such automatic and quantitative measurements are critical for aiding current and future diagnosis. We believe one of the greatest impacts of our method is that it gives the clinician quantifiable data for this task and may allow readjustment of the rigid scoring intervals provided in the AOSI. With a study on a larger population, new time intervals (and their variability) for scoring may be discovered, and these false positives could be analyzed not as a strict “pass” or “delayed” but as something in between.

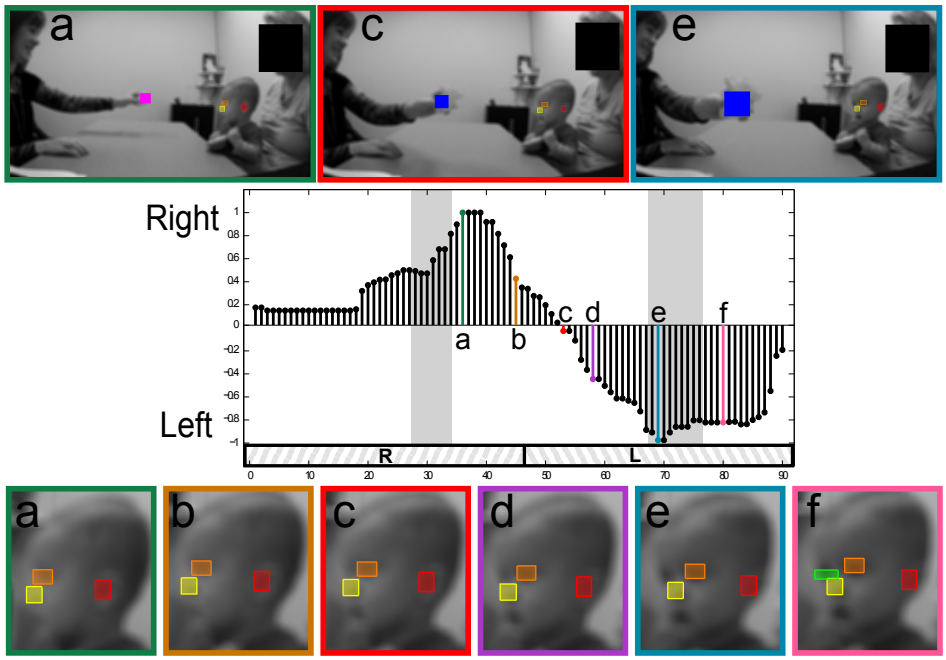
Table 2.3: Results of Visual Tracking task. Scores from the clinician (Clin.), automatic method (Automatic), a psychiatrist (Psy.), and two students (St. 1 and St. 2) for each trial. A trial can be considered "passed" (Pass), "delayed" (Del), "interrupted" (Int), "partial" (Prt), or "no tracking" depending on how smoothly the child visually tracks the object (colors are added to facilitate the comparison).

Part.	First Trial Score					Second Trial Score				
	Clin.	Automatic	Psy.	St. 1	St. 2	Clin.	Automatic	Psy.	St. 1	St. 2
#1	Pass	Pass	Pass	Pass	Pass	Pass	Int	Int	Int	Int
#2	Int	Int	Pass	Int	Pass	Int	Int	Pass	Pass	Pass
#3	Del	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
#4	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
#5	Prt	Int	Pass	Int	Pass	Prt	Prt	Del	Prt	Prt
#6	Pass	Pass	Pass	Pass	Pass	-	-	-	-	-
#7	-	-	-	-	-	Int	Int	Int	Prt	Prt
#8	Pass	Pass	Pass	Del	Pass	Pass	Pass	Pass	Pass	Pass
#9	Pass	Pass	Pass	Del	Pass	Pass	Pass	Pass	Pass	Pass
#10	Pass	Pass	Pass	Pass	Pass	Int	Int	Int	Prt	Prt
#11	Pass	Pass	Pass	Pass	Pass	Int	Int	Int	Prt	Prt
#12	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass

2.4.2 Visual tracking

Table 2.3 summarizes the results of our method, the clinical assessments, and the ratings by a child/adolescent psychiatrist and two psychology students for the Visual Tracking task. As in Table 2.2, the trials with blank spaces and a horizontal line could not be used by our automatic method. The simple output of our method allows a clinician to easily assess each trial by visual inspection and score the trials as either "pass," "interrupted," "partial," or "no tracking."

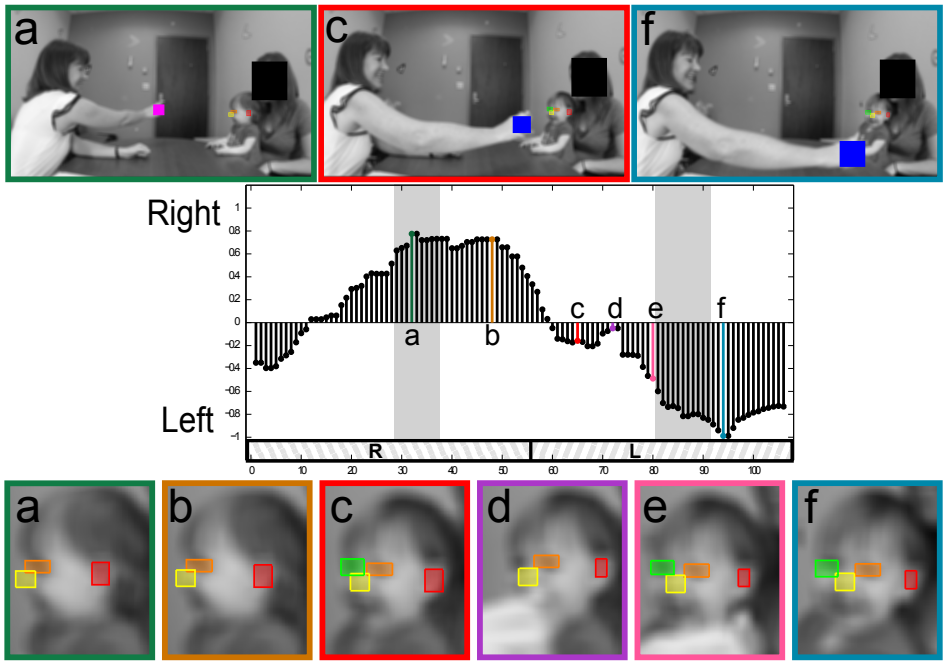
Out of the 14 trials that the clinican assessed as "pass", our method agreed with 13 of them and scored an "interrupted" for 1 of the trials. For all the 4 trials the clinician assessed as "interrupted," our automatic method was in agreement. The clinician scored two trials as "partial," our method scored one of them as "partial" and the other as "interrupted."



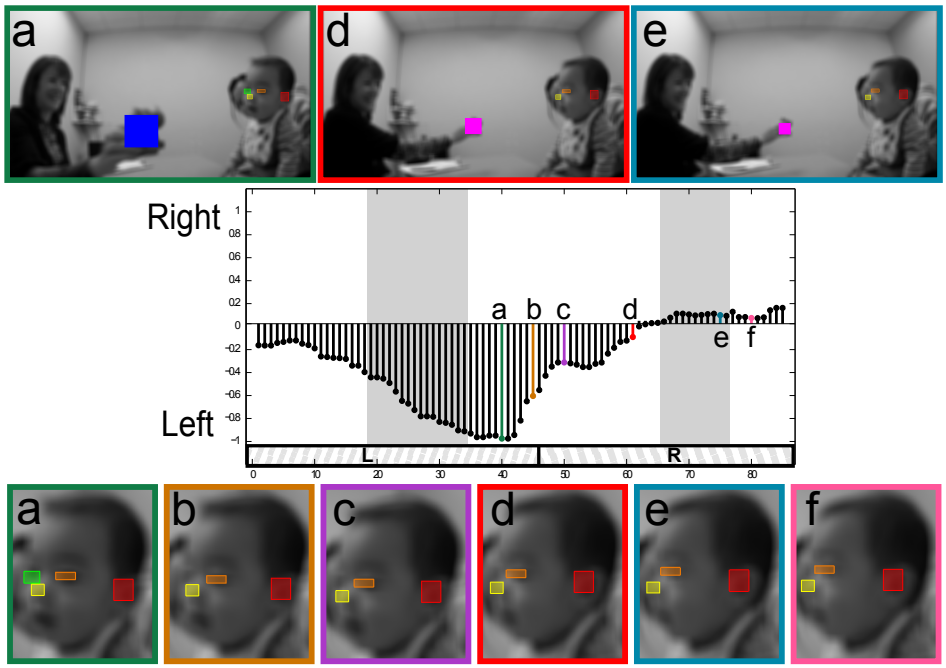
(a)

Figure 2.5: First example of the Visual Tracking task. **Top:** the clinician holding the object, when the object is at one extreme side (right or left), and when the object is at the other extreme side. **Middle:** changes in the yaw motion ($\widehat{\text{yaw}}$ values in the y-axis) for every frame (x-axis). The boxes labeled 'R' and 'L' represent when the object is to the right and left of the participant respectively. The gray shaded areas represent when the object is not moving and at an extreme side (either right or left). **Bottom:** 6 examples of the infant's face during the task. Colors and letters identify corresponding images and spikes in the graph.

Figure 2.5 shows important examples of our results for the Visual Tracking task. Example 2.5(a) demonstrates a participant that received a "passed" on the AOSI's Visual Tracking task, since the participant was able to smoothly track the object with minimal delay as the object approached the participant's right. In example 2.5(b), the participant exhibited "interrupted" tracking motion. The participant's tracking of the object was interrupted as the object moved across the clinician's face. Instead of tracking the object as it moved across the clinician's face, the participant stopped tracking the object and looked at the clinician for 0.46s (14 frames) before continuing to track the object as it moved to the participant's left. Such short behaviors can only be detected by an



(b)



(c)

Figure 2.5: (Continued) Two examples of the Visual Tracking task.

automatic sytem. In example 2.5(c), the participant displays a “partial” tracking score on the AOSI. As the object crosses the clinician’s face, the participant completely stops tracking the object and instead looks straight at the clinician.

2.4.3 Comparision with non-experts in ASD

Table 2.4: Number of agreements with the Autism expert for each participant in the two visual attention tasks and overall inter-rater reliability using weighted Cohen’s kappas. See tables 2.2, 2.3

Task	Trials	Automatic	Psychiatrist	Student 1	Student 2
Disengagment	27	**25**	22	13	16
Tracking	22	**19**	16	13	14
Total	49	**44**	38	26	30
Inter-Rater Score	-	**0.75**	0.37	0.27	0.27

The results obtained by the child/adolescent psychiatrist and two psychology students are presented in tables 2.2 and 2.3. Out of the 27 Visual Disengagement trials (Table 2.2), the two psychology students agreed with the clinician on 13 and 16 of the trials respectively, while the child/adolescent psychiatrist agreed on 22 trials. Our system agreed in 25 cases. Similarly for the 22 Visual Tracking trials (Table 2.3), the two psychology students agreed with the clinician on 13 and 14 of the trials respectively, while the child/adolescent psychiatrist agreed on 16 trials. Our system agreed on 19 cases. We also provide the inter-rater reliability score of each individual compared with the clinician. For this, we use weighted Cohen’s kappas with a range of 0-1, where 1 means complete agreement. Table 2.4 provides a summary of these results. Our method’s accuracy is highlighted, when compared against the results obtained by the child/adolescent psychiatrist and two psychology students, it clearly outperforms non-expert assessments in both accuracy and inter-reliability. Note that while not experts in ASD, the training of these testers is significantly above what is standard in developing

regions for example. These results not only illustrate the human training that needs to be done for these visual attention tasks but also stress the novelty of our method and its quantitative results.

2.4.4 The ball playing activity

As stated before, the clinician's assessment of shared interest is partially done in the ball playing activity. Since the AOSI assessment of shared interest is done on a holistic scale, we provide examples of where our automatic method could prove useful. Figure 2.6 shows examples of our results from the ball playing activity during the freeplay session. For this particular activity, the clinician rolls a ball to the participant and analyzes if the participant shows shared interest. According to the AOSI, the participant shows shared interest if he/she either looks at the clinician or his/her caregiver after receiving the ball. Our automatic method is able to record and display the changes in the pitch motion of the participant. This allows the clinician to not only determine if the participant looked up after receiving the ball, but also how long it took him/her to look up and how long he/she became fixated on the ball. For both provided examples, the participants looked back up at the clinician after receiving the ball. In the first example, the participant looked up at the clinician within 0.73s (22 frames) of receiving the ball. On the other hand, it took the participant in the second example 7.17s (251 frames) to look up at the clinician after receiving the ball. Although each participant showed shared interest, the participant in the second example looks at the ball nearly 6.5s longer than the first participant before he/she looks back up at the clinician. Not only does our method estimate the participants' pitch motions, but it also provides other information such as the time it takes for each participant to look up after receiving the ball. The automatic detection of this time difference is an example of our algorithm's potential impact on the research community.

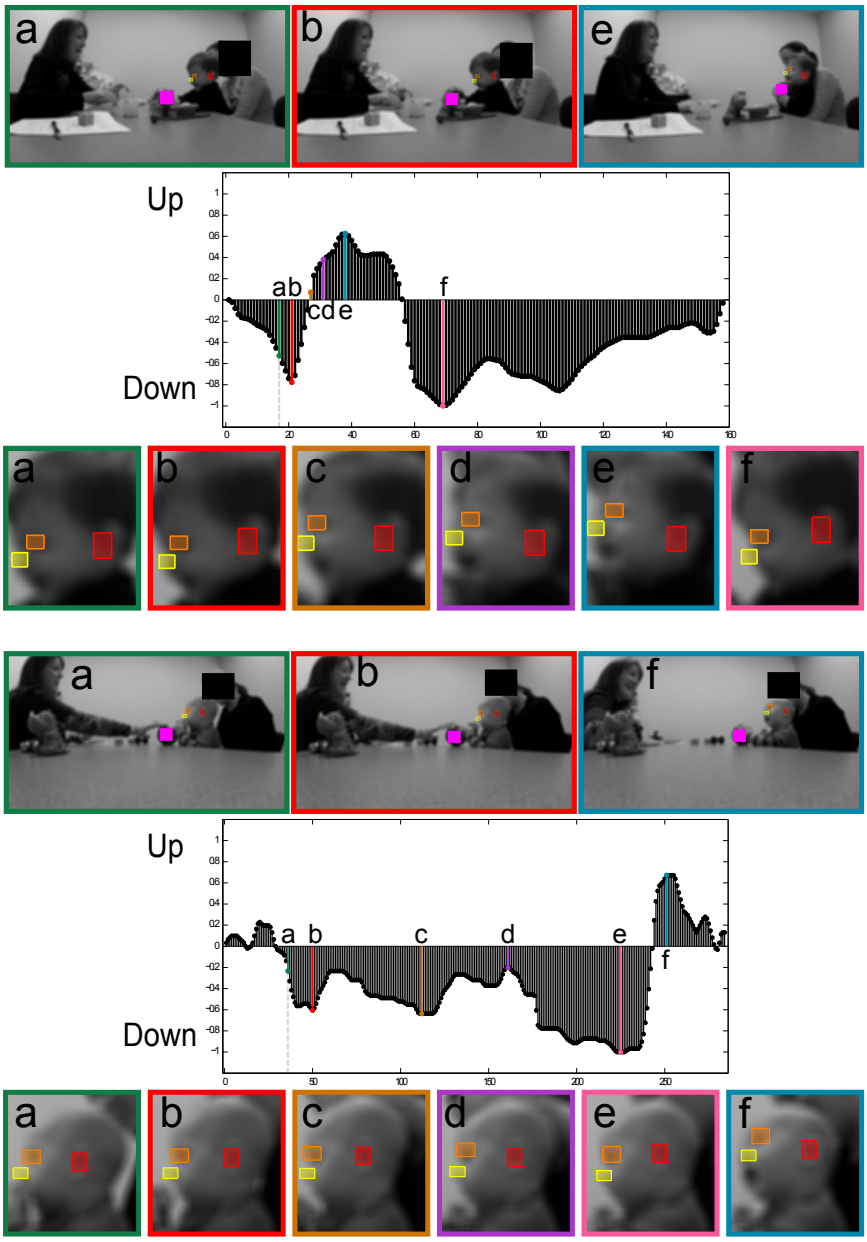


Figure 2.6: Two examples of the Ball Playing activity. **Top:** when the ball contacts the child, when the child looks down at the ball, and when the child looks up at the clinician. **Middle:** changes in the pitch motion (y-axis) for each frame (x-axis). The dotted line represents when the ball contacts the participant. **Bottom:** 6 examples of the infant's face during the administration. All facial features are automatically detected and tracked. Colors and letters identify corresponding images and spikes in the graph.

2.5 Conclusions

In this chapter, we developed a head motion estimation method for critical AOSI tasks and activities, and presented results comparing our method with that of expert and non-expert clinicians. With the goal of aiding and augmenting the visual analysis capabilities in evaluation and developmental monitoring of ASD, we proposed (semi-)automatic computer vision tools to observe specific important behaviors related to ASD elicited during AOSI, providing both new challenges and opportunities in video analysis. The proposed tools significantly reduce the effort to only requiring interactive initialization in a single frame, and liberates the potential casual examiner from extensive training, adding both accuracy and objectivity. The benefits of the results obtained with our automatic method for head motion estimation are threefold. First, it provides accurate quantitative measurements for the AOSI tasks, improving the shareability of clinical records while not compromising anonymity. Second, it can also prove beneficial in the discovery of new behavioral patterns by easily collecting large amounts of data and using data mining on them. Third, it increases the granularity of the analysis by providing results at a finer scale.

3 Elliptical cell segmentation and analysis of yeast cell clusters

3.1 Chapter summary

In this chapter, we develop computer vision tools to automatically analyze yeast cell clusters from differential inference contrast (DIC) microscopy and fluorescence images. The main motivation for this work is to provide biologist with robust and high throughput tools to aid in yeast cell analysis. More specifically, our tools automatically segment corresponding cell and fluorescence images and compute a multitude of cell and fluorescence properties including: area, location, eccentricity, and fluorescence intensity. We present a novel extension to the circular Hough transform (CHT) to segment circular and elliptical cells, as well as present segmentation methods for three types of fluorescence labeling. We present results demonstrating our cell segmentation's versatility by segmenting two popular yeast strains: Candida albicans and Saccharomyces cerevisiae. We also present results demonstrating the robustness of our fluorescence segmentation by segmenting five different fluorescence markers. Lastly, we validate our developed program as well as demonstrate practical applications for the yeast cell community. The work in this chapter is cited from [17, 16, 15].

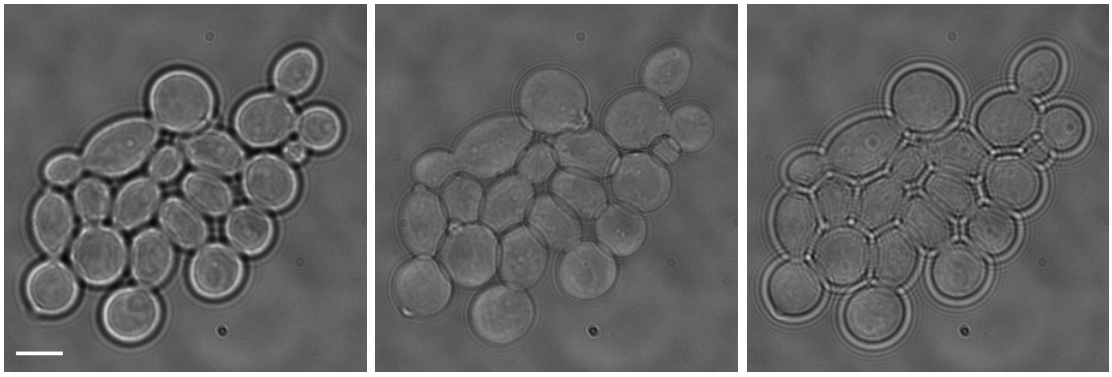


Figure 3.1: Example of the sensitivity of DIC microscopy images of the same cluster. The focal depth of the middle image has only a 28μm difference between the other two images. The scale bar represents 5μm.

3.2 Introduction and motivation

Automatic segmentation and analysis of microbiological cells from differential inference contrast (DIC) microscopy images has greatly impacted the microbiology community and has been a widely studied topic [29, 4, 48]. It allows for high throughput analysis, creating new opportunities such as analyzing cell interactions in 3D [27]. Recently a lot of focus has been on discovering pathways and manipulating responses of Candida albicans (C.albicans), a widely studied opportunistic fungal pathogen, and Saccharomyces cerevisiae (S.cerevisiae), a model eukaryotic organism, yeast strains[24, 34]. Cells from these strains exhibit three main morphological structures: blastospores (circular), psuedohyphae (elliptical), and hyphae (long and very thin). A widely interesting area has been pseudohyphal growth and its association with virulence in C.albicans clusters [45, 5].

Pseudohyphal growth can be captured from DIC microscopy, a widely accepted method to image transparent cells such as the mentioned yeast strains [21]. Although DIC microscopy can yield high spatial resolution, its depth of focus is extremely sensitive. Due to this sensitivity, when imaging a cluster of yeast cells, it can be a tedious task to ensure

all of the cells are in focus. For the cells that are not completely in focus, halo artifacts appear around them (Figure 3.1); furthermore, when cells are completely in focus, intracellular structures are imaged. For cell segmentation purposes, intracellular structures imaged in DIC images are considered noise; however, they are of great importance when labeled with specific fluorescent proteins (FPs).

Intracellular properties, such as protein expression and localization, can be examined via FPs. FPs are highly specific and allow for monitoring of protein expressions in living cells; notably, FPs fused to different proteins of interest and excitation wavelengths can be used simultaneously, allowing for monitoring of multiple intracellular structures. Three important FPs that provide insight to spindle pole body, sub-nuclear, and nuclear phenomenon in yeast strains are: tub-4, nop-1, and hhf-1 respectively. More specifically, the fluorescence intensity, locations, and number of segmentations per cell of hhf-1 and nop-1 segmentations directly correlate with DNA content and cell cycle regulation; while the number of tub-4 segmentations per cell only correlates with DNA replication. The expression of these FPs greatly differ (see second row of Figure 3.6 for distinction between fluorescence types).

We propose to develop a user-friendly program that accurately segments cell and fluorescence signals from both DIC and fluorescence images respectively. Cell segmentation relies heavily on finding the outside edges of cells in images and its performance can be impaired from intracellular noise. Two popular techniques adapted into cell segmentation methods are the circular Hough transform (CHT) [10] and active contours [26]. The CHT presents an efficient and accurate solution to cell segmentation [14], but the main drawback is its restriction to only segmenting blastospores. On the other hand, active contours allow for segmentation of more diverse shapes [2] such as pseudohyphal formations; however, these algorithms are computationally expensive and imaged intracellular structures can pose problems. To ensure there are minimal issues regarding

intracellular noise, we propose to use multiple DIC images that are captured at different focal depths. This not only ensures that between all of them, we are receiving an accurate representation of the entire cluster, but it also does not restrict the user to have to determine an ideal focal depth to image an entire cluster at.

For our fluorescence segmentation methods, we propose to classify hhf-1 as a fluorescence marker with a high signal-to-noise ratio (SNR), nop-1 as having a medium SNR, and tub-4 as having dim labeling. Both hhf-1 and nop-1 expressions can be seen on all of the images captured at different focal depths, while tub-4 expression is dimly scattered throughout parts of the focal depth images.

The work here is aimed at creating an efficient and accurate yeast cell analysis program that can handle both C.albicans and S.cerevisiae clusters from DIC microscopy images, as well as corresponding fluorescence images. We first present our elliptical cell segmentation algorithm which uses both the CHT and a simplified active contour method. Then, we present fluorescence segmentation algorithms that handle three different classifications of fluorescence labeling. We demonstrate that by combining cell and fluorescence segmentations together, we are able to create a user-friendly program which the yeast cell community can apply to a multitude of applications. And lastly, we present experimental results and provide applications which validate our method.

3.3 Proposed cell analysis method

The proposed framework is separated into two sections: cell segmentation and fluorescence segmentation. Our cell segmentation method utilizes vector-valued edges, the circular hough transform (CHT), as well as our implemented extended-CHT. Our fluorescence segmentation is done in a manner that is independent of the cell segmentation‘s results, and its primary components are Otsu‘s thresholding method and the CHT.

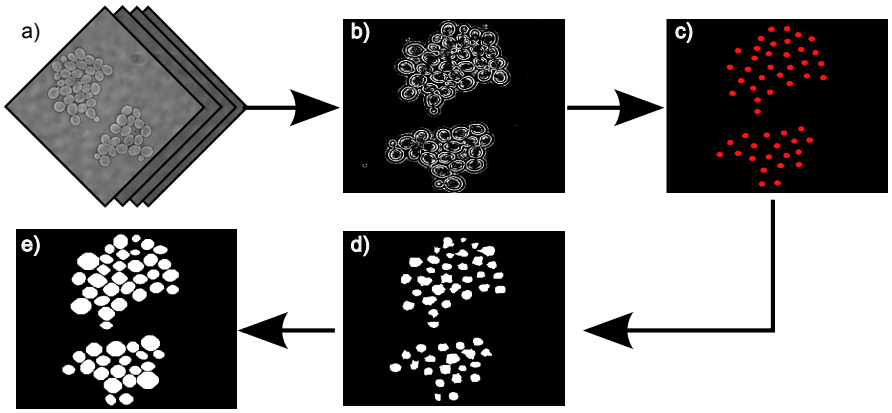


Figure 3.2: Overview of proposed cell segmentation method. **a)** input multiple images of same cluster(s) at different focal depths. **b)** result of vector-valued edges. **c)** possible cell locations detected from CHT. **d)** cell shapes as a result of extended-CHT. **e)** final cell segmentation after fitting and extending ellipses to estimated cell shapes.

3.3.1 Cell segmentation

Vector-valued edges

We assume that as an input, we receive multiple images of the same cluster imaged at different focal depths. To determine similar edges throughout all of the images, we follow a technique of determining the eigenvalues in multispectral images [8]. Let $\mathbf{I}(x_1, x_2)$ be a vector valued image with components $I_n(x_1, x_2), n = 1, ...N$, where N represents the number of input images per cluster; and in our case, we convert each image to grayscale. We are interested in finding pixels with high rates of change, otherwise known as edges, thus we calculate the squared Euclidean norm. Let us denote $g_{ij} := [\frac{d\mathbf{I}}{dx_i} \cdot \frac{d\mathbf{I}}{dx_j}]$, then the squared norm is given by:

$$d\mathbf{I}^2 = \begin{bmatrix} dx_1 \\ dx_2 \end{bmatrix}^T \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \begin{bmatrix} dx_1 \\ dx_2 \end{bmatrix} \tag{3.1}$$

This quadratic form is referred to as the first fundamental form. The eigenvalues of

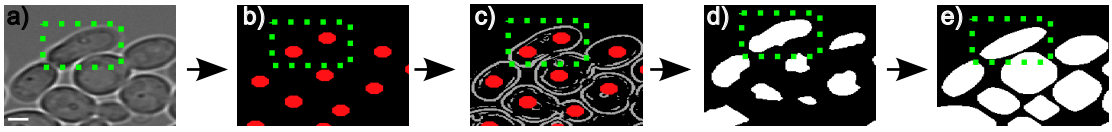


Figure 3.3: Example of our ellipse and circle detection. **a)** one of the input DIC images. **b)** CHT result. **c)** CHT result overlaid on prominent edges provided from λ_+. **d)** Result after extending CHT detections to nearest prominent edges. **e)** final cell segmentations after fitting ellipses to results from **d**. The dotted green box encompasses an example of elliptical segmentation. The scale bar represents $5\mu m$.

the middle matrix, g_{ij}, correspond to the magnitudes of the gradients present throughout all of the input images. For this application, we are interested in the eigenvalue with the maximal magnitude:

$$\lambda_+ = \frac{g_{11} + g_{22} + \sqrt{(g_{11} - g_{22})^2 + 4g_{12}^2}}{2} \tag{3.2}$$

The resulting matrix λ_+ represents the maximal rate of change per pixel throughout all the input images; in other words, we acquire a single matrix who‘s values represent how well-represented an edge is throughout all the input images (see Figure 3.2**b**).

Extended-CHT

Ellipse detection involves finding five unknown parameters: the vertical and horizontal locations of the center, the orientation, and the minor and major axes. Estimating these five parameters involve huge accumulators, creating extremely computationally expensive and inefficient estimations. On the other had, the CHT detects circles and relies on three parameters: the radius, and the vertical and horizontal locations of the center. By predefining the radii range, the process of detecting circles reduces to finding only the center locations of possible detections. Our novel approach is based off finding the two end points of an ellipse via the CHT and then fitting the ellipse by extending the edges of the detected circles until they reach prominent edges provided by the applying Otsu's

thresholding method to the vector-valued edge result. Note that although we automatically determine the prominent edge-threshold via Otsu's method, we allow the user to adjust this threshold (see Figure 3.5 in the Results).

We employ a fast and efficient implementation of the CHT [38] which requires a predefined radii range for circle detection. By predefining the radii range in the CHT, the search can be reduced to only 2 dimensions; the objective becomes finding the centers of the circle. Let x_0 and y_0 be the centers we are trying to determine, and R be a predefined radius. We wish to determine x_0 and y_0 such that:

$$x = x_0 + Rcos(\theta)$$

$$y = y_0 + Rsin(\theta)$$

is well represented in the input image. This can be efficiently solved by a Hough accumulator array with only two unknown parameters, the center locations x_0 and y_0. Where the parameter space is the same as the input image, and each entry in the array represents the voting that a circle's center is located there. A high voting in a given pixel, represents a strong chance of a circle's center location.

Using the vector valued edge result, λ_+ , as an input image to the CHT, we apply a lenient threshold on the resulting accumulator matrix to detect all possible locations of circular cells and possible ends of elliptical cells (see figures 3.2**c** and 3.3**b**). We then overlay the CHT detections on top of the λ_+ result (Figure 3.3**c**). Following an approach similar to active contours, we extend the edges of the detections until they touch prominent edges provided by λ_+. From this, we are not only able to combine detections into elliptical segmentations but also acquire the following information regarding each cell‘s elliptical shape: the vertical and horizontal center locations, the orientations, and approximate ratio between the cells‘ major and minor axes (see Figure 3.3**d**).

Lastly, we fit ellipses to each segmentation provided by the results from our extended-CHT. Our extended-CHT tends to segment the intracellular area rather than the entire cell (e.g. see segmentation to the right of the green box in Figure 3.3**d**). To ensure the final segmentation is not just a representation of intracellular area, we utilize the fact that a cell in a cluster has budded from another cell. Meaning that every cell has at least one neighboring cell that nearly share the same edge. Thus, after fitting ellipses to the intracellular area, we dilate the fitted ellipses in a given cluster until they are nearly touching; this is our final segmentation result (see figures 3.2**e** and 3.3**e**).

3.3.2 Fluorescence segmentation

To accommodate for multiple types of fluorescence labeling, we developed three different fluorescence segmentation algorithms: high SNR, medium SNR, and dim labeling. Similar to our cell segmentation algorithm above, our fluorescence segmentation algorithms all utilize multiple input images at different focal depths. To segment the fluorescence signals in the high and medium SNR cases, we first employ vector-valued edges to create a single image containing edges that are present throughout all of the input images. Then using the widely accepted Otsu‘s thresholding method [37], we are able to accurately and efficiently segment the fluorescence signals.

In the dim labeling case, a fluorescence signal is not captured at different focal depths. Meaning that at each focal depth image, a different population of signals are captured. These fluorescence signals are also very small compared to the high and medium SNR labeling. The dim labeling signals present circular shaped expressions with radii between 2 to 4 pixels. Instead of using vector-valued edges like the previous proposed approaches, our dim labeling segmentation deals with each focal depth image independently. We employ the same CHT implementation as in 3.3.1 but restrict the radii search

range to 2 to 4 pixels. After segmenting each image, we merge these segmentations together to create a final segmentation image.

3.3.3 Combining cell and fluorescence segmentations

Lastly, we compute multiple properties for the cell and fluorescence segmentations as well as provide the results in a user friendly manner. By employing the connected components algorithm [41] for both the final cell and fluorescence segmentations, we are able to compute individual segmentation properties including cell and fluorescence area, location, and eccentricity. Based on individual cell and fluorescence segmentations‘ locations, we are able to correlate which fluorescence segmentations lie within which cell segmentations.

Another computation our method provides is fluorescence intensity. The analysis of fluorescence intensity is done in a normalized and accurate fashion in the sense that the computation of the intensities is not altered between different imaged clusters. Our algorithm computes fluorescence intensity (Int) of the j^{th} fluorescence segment by accumulating the total fluorescence intensity within the segment‘s mask across all of the input images:

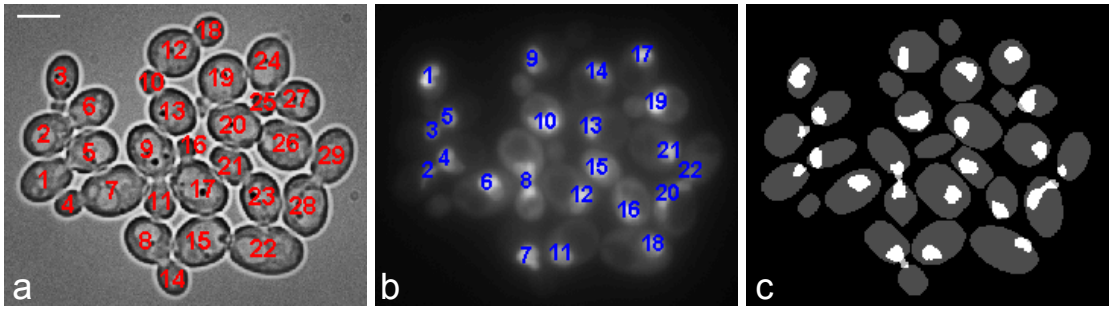
$$Int_j = \sum_{k=1}^{N} Input_k * Mask_j \tag{3.3}$$

Let $Input_k$ represent the k^{th} input fluorescence image where $k = 1, ..., N$ and N is the total number of focal depths for the fluorescence images. Let $Mask$ be a binary image where $Mask_j$ represents the j^{th} fluorescence segmentation; meaning the entire image has the value 0 except the area containing the j^{th} labeled fluorescence segmentation which has the value 1. Thus the morphological multiplication of $Input_k$ with $Mask_j$ can be understood as discarding the entire $Input_k$ image except the area that

corresponds to $Mask_j$. One must be careful to ensure that none of the original fluorescence images reach saturation. Saturated images can distort the fluorescence intensity calculation since pixel saturation reduces the accuracy of the underlying fluorescence intensity distribution.

The correlations between the cell and fluorescence segmentations as well as their properties are organized in an Excel spreadsheet. We also output and save images of the segmentations, as a Matlab file, so users are able to double-check and visually see the results. An example of our method‘s output is shown in Figure 3.4. The organization of the spreadsheet (Figure 3.4**d**) is as follows. The first column corresponds to the number of the imaged cluster and in column 2, each individual cell segmentation receives a numbered label (the numbers are assigned via the connected components algorithm). Columns 3-5 provide the cell segmentations‘ computed properties including the number of fluorescent segmentations within each cell segmentation, cell area in terms of pixels, and cell eccentricity. Column 6 shows the number of the labeled fluorescent segmentation that lies within the labeled cell segmentation in the same row. While columns 7-10 provide the fluorescent segmentations‘ computed properties including the number of cell segmentations the fluorescent segmentation touches, fluorescent intensity, the fluorescent segment‘s area in terms of pixels, and the segment‘s eccentricity. In instances where a cell segmentation does not have any fluorescent segmentations within it, columns 6-10 will be zeros (see Cell #‘s 4, 10, 14, 16, 18, and 25 in Figure 3.4). If a given fluorescent segmentation touches more than one cell segmentation, the specific fluorescent segmentation‘s properties are repeated for each cell segmentation it touches (see Cell #‘s 9 and 11 and GFP # 8 in Figure 3.4). When a given cell segmentation has more than one fluorescent segmentation within it, more rows are added with the specific cell‘s properties and the other fluorescent segmentations‘ properties.

To visually see the labeled segmentation results, our method provides a *.mat* file that



d

Field #	Cell #	# of GFP	Cell Area	Cell Ecc	GFP #	# of Cells	Intensity Vol	GFP Area	GFP Ecc
1	1	1	21.77	0.78	2	1	0.57	2.01	0.73
1	2	1	22.98	0.57	3	1	0.48	1.75	0.69
1	3	1	15.73	0.71	1	1	2.33	5.46	0.76
1	5	1	17.34	0.89	4	1	1.35	3.70	0.76
1	6	1	17.96	0.71	5	1	1.20	3.64	0.55
1	7	1	30.70	0.81	6	1	2.18	4.59	0.71
1	8	1	23.18	0.73	7	1	2.00	4.92	0.78
1	9	1	23.93	0.78	8	2	2.84	6.58	0.94
1	11	1	18.40	0.61	8	2	2.84	6.58	0.94
1	12	1	29.16	0.57	9	1	1.25	3.71	0.88
⋮									
1	4	0	6.67	0.65	0	0	0.00	0.00	0.00
1	10	0	8.19	0.65	0	0	0.00	0.00	0.00
1	14	0	11.98	0.53	0	0	0.00	0.00	0.00
1	16	0	12.85	0.87	0	0	0.00	0.00	0.00
1	18	0	7.57	0.46	0	0	0.00	0.00	0.00
1	25	0	8.37	0.53	0	0	0.00	0.00	0.00

Figure 3.4: Examples of automatic segmentation outputs (a-c) and spreadsheet output (d) from a C. Albicans cluster labeled with eno1. **a)** An input DIC image overlaid with cell segmentation labels. **b)** An input fluorescence image overlaid with fluorescent segmentation labels. **c)** Resulting cell and fluorescent segmentations overlaid on one another. The darker gray ellipses represent the cell segmentations and the white segmentations represent the fluorescent segmentations. In examples a-c, the red numbers correspond to cell number labels while the blue numbers correspond to fluorescent number labels. The scale bar in **a)** represents $5\mu m$.

can be called on with the provided $show-images.m$ function. With this function, one specifies which imaged cluster number he/she wishes to display the results from. The function displays three images: one that shows the cell segmentation overlaid on top of one of the original DIC images; an image showing the fluorescent segmentation overlaid on top of one of the original fluorescent images; and an image showing both the cell and fluorescent segmentation results overlaid on top of each other (see Figure 3.4**a-c**). For all three images, the cell and/or fluorescent segmentations are individually numbered in a manner that corresponds with the numbered labeling in the results spreadsheet.

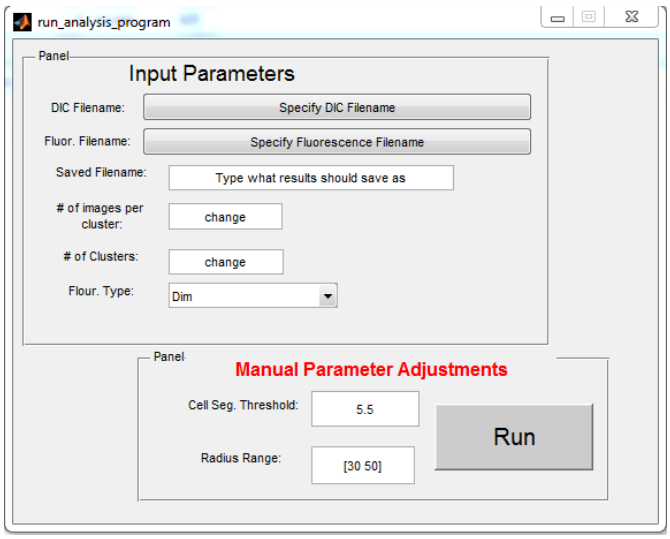


Figure 3.5: Example of graphical user interface for our cell analysis program. The user inputs the desired filenames, number of images per cluster, the number of clusters, and the fluorescence type. We also allow for manual adjustments to our prominent edge-thresholding parameter and the CHT radii range for cell detection.

3.4 Experimental Results

Although our method is intended for automatic analysis of cell clusters, we implemented an easy-to-use graphical user interface with adjustable parameters for prominent edge-thresholding and radii range so users are able to fine tune these parameters for their needs (see Figure 3.5). The duration of analysis largely depends on the number of cells in the cluster and the computer hardware. The approximate duration of the analysis program on a cluster of 30 cells using an Intel i7 3.4 GHz/16 GB desktop running 64-bit Windows 7 is in the order of a minute. We first show examples of our cell and fluorescence segmentation results. Then we provide two example applications of our program. The first example demonstrates the validity of our fluorescence intensity calculation. While, the second example demonstrates a possible application of our program with the analysis of untreated and treated cells.

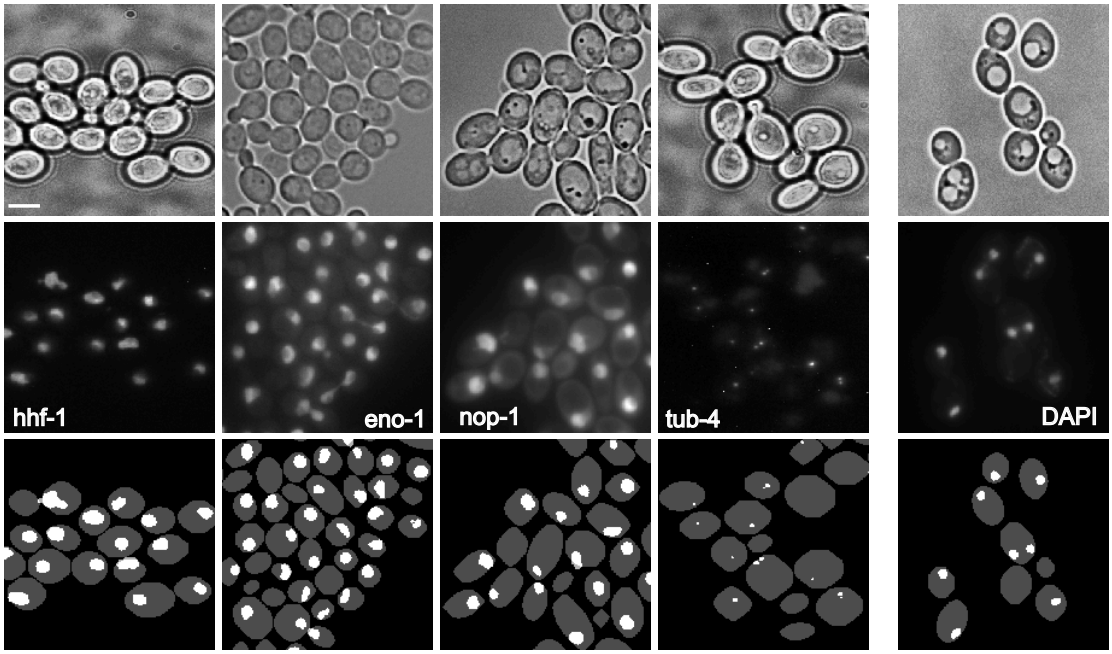


Figure 3.6: Examples of cell and fluorescent segmentations compared with their original images using C. Albicans (first four columns) and S. Cerevasiea (last column). The first and second rows consist of original DIC microscopy and fluorescence labeled images respectively. The third row displays the cell and fluorescent segmentations overlaid together where the gray and white segmentations represent cell and fluorescent segmentations respectively. hhf-1 and DAPI are examples of high SNR labeling. eno-1 and nop-1 represent medium SNR. And tub-4 is an example of dim labeling. The scale bar represents $5\mu m$.

3.4.1 Segmentation examples

Our method was originally developed for high throughput analysis of morphological and intracellular properties of C. Albicans using hhf-1, eno-1, nop-1, tub-4 fluorescent labeling [16]; however, we have modified and tested our method to ensure it can also be used with S. Cerevasiea cells and more generalized fluorescent labeling. To demonstrate the versatility of our method, we provide example segmentation results of both yeast strains as well as five different fluorescence markers: hhf-1, eno-1, nop-1, tub-4, and DAPI(Figure 3.6). Our program requires an input of at least three DIC images and fluorescence images of a cluster at different focal depths. In our setup, we usually image

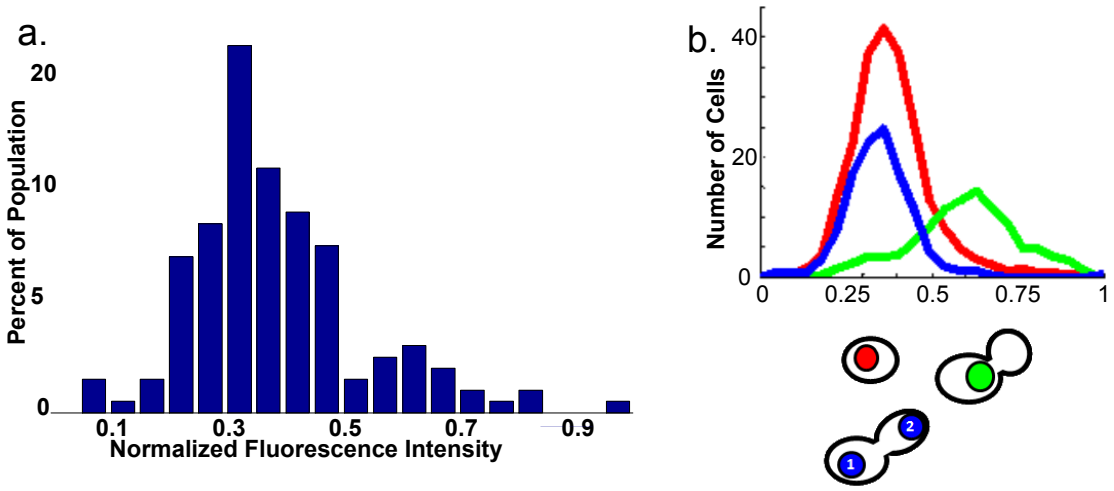


Figure 3.7: Normalized fluorescence intensity distributions of C.albicans labeled with hhf-1 FPs. **a)** bimodal fluorescence intensity distribution, representing segmentations with replicated and unreplicated DNA content. **b)** fluorescence intensity distributions of segmentations after they are categorized based on the image below. Red, green, and blue colors correspond to fluorescence segmentations belonging to a single cell, a budding cell, or a completed budding cell respectively.

clusters anywhere from 12 to 21 different focal depths with depth intervals at 2 to 4μm.

3.4.2 Fluorescence intensity distributions

Histones are the primary protein in chromatin and are responsible for packaging DNA. Histone replication occurs alongside DNA replication during the S phase of the cell cycle. We tested our fluorescence intensity calculations by labeling clusters of C.albicans with a histone specific FPs, hhf-1, and ran our program to verify the fluorescence intensity calculations. We expect to see a bimodal fluorescence intensity distribution, where the first peak would represent segmentations with unreplicated DNA content, and the second peak would represent segmentations with replicated DNA content and would have roughly twice the amount of fluorescence intensity as the first peak. We ran two experiments and the results are shown in Figure 3.7. In the first experiment, our normalized fluorescence intensity distribution displayed a bimodal distribution where the first peak

was roughly 0.325 and the second peak appearing at 0.625 (Figure 3.7**a**). We took the results one step further and categorized fluorescence segmentations as belonging to a single cell, a budding cell, or a completed budding cell (see bottom picture of Figure 3.7**b** for categorical distinctions). We expected the fluorescence segmentations belonging to the single cell and completed budding cell categories to have unreplicated DNA content and exhibit a similar intensity distribution between them; while we expected the segmentations belonging to budding cells to have the replicated DNA content and display a distribution centered around a value that is twice that of the other categories. After creating a fluorescence intensity distribution for each of these three categories, our results proved the validity of our fluorescence intensity calculations (Figure 3.7**b**).

3.4.3 Cell and fluorescence segmentation properties

Fluconazole is an well established antifungal agent for the localization of C.albicans infections; however, due to its fungistatic qualities, resistant strains of C.albicans can arise and become predominant [33, 15]. Our cell and fluorescence segmentation method was used to monitor the morphological and intracellular responses of fluconazole-treated and untreated C.albicans cells labeled with nop-1 FPs at 3 hour time intervals. We incorporated our method to show that the average compartment area and variability significantly increased after 6 hours of fluconazole exposure. Another result we were also able to depict was that large compartments with 3 nop-1 masses started appearing 9 hours after initial exposure. We also saw a correlation between compartment size and the number of nop-1 masses that the compartment contains (see [16, 15] for more experimental details, results, and discussion).

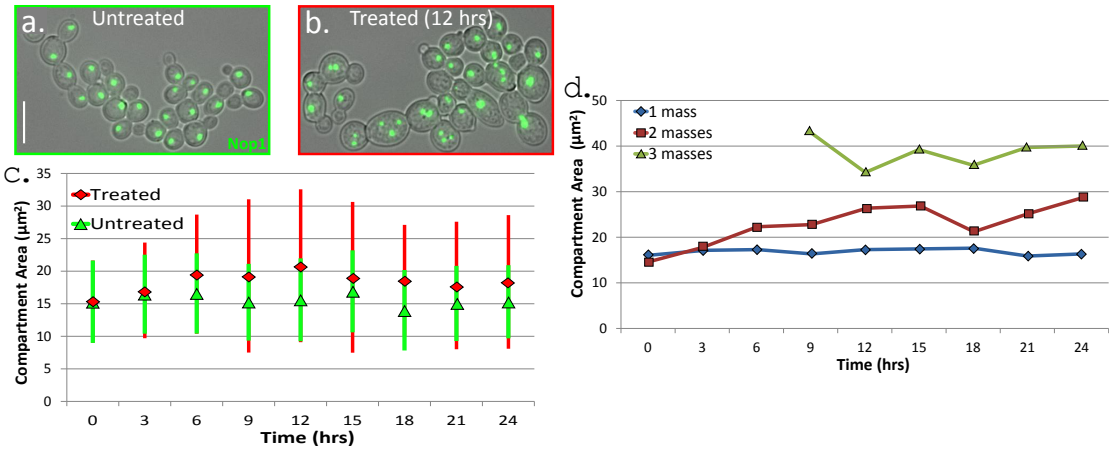


Figure 3.8: Examples of compartment size variability and intracellular changes after treatment of fluconazole in C.albicans clusters labeled with nop-1 FPs. **a)** an example image of an untreated cluster. **b)** an example image of a cluster 12 hours after fluconazole treatment. **c)** compartment area increases and varies more after fluconazole treatment. Large abnormal compartments appear after 6 hours of exposure. **d)** increase in compartment size correlates with an increase in the number of nop-1 masses that the compartment contains. Scale bar represents $5\mu m$.

3.5 Conclusions

In this chapter, we developed a program to analyze cell clusters from two prominent yeast strains: C.albicans and S.cerevisiae. The novelty of the presented method involves extending the CHT to accommodate for elliptical cells. We also present three fluorescent segmentations, and show segmentation results demonstrating the versatility of our algorithms. Lastly, we present two example applications of our developed program. The first example involved creating fluorescence intensity distributions to distinguish between segmentations containing replicated and unreplicated DNA. We also validated the underlying fluorescence intensity distributions by categorizing the segmentations as belonging to single cells, budding cells, and completed budding cells; and then showing the underlying intensity distributions are what is expected. The fluorescence segmentations belonging to the single and completed budding cells displaying similar intensity distributions whose means were nearly half of that calculated by the budding cells. The

fluorescence segmentations belonging to the budding cells exhibited an intensity distribution that indicated it contained replicated DNA content. Lastly, we present examples of using both the cell and fluorescence segmentations to demonstrate morphological and intracellular changes of C.albicans clusters tagged with nop-1 after exposure to fluconazole.

4 Conclusions

The main conclusion of this thesis is that, computer vision approaches can aid and improve a variety types of analysis including clinical assessment of behavioral disorders and quantification of yeast cells. These developed computer vision tools can allow for high throughput analysis of these applications; allowing for new information to be brought to light through data mining techniques.

Specifically, we have shown that we are able to estimate head motions during autism assessment tasks and activities in a non-intrusive manner. We provided results demonstrating our tools can aid the practitioner's assessment in an objective manner as well as provide new information such as head motion patterns.

We also have shown that we are able to accurately and automatically quantify yeast cell and fluorescence images. We demonstrate the versatility of our method by providing multiple segmentation examples. The possible applications of our developed yeast cell analysis program are endless. We present two application: creating fluorescence intensity distributions, and monitoring morphological and intracellular properties of C.albicans clusters treated with fluconazole and non treated.

5 Bibliography

[1] S. Basu, I. Essa, and A. Pentland. Motion regularization for model-based head tracking. In *ICPR*, 1996.

[2] K. Bredies and H. Wolinski. An active-contour based algorithm for the automated segmenation of dense yeast populations on transmission microscopy images. *Computing and Visualization in Science*, 2012.

[3] S. Bryson, L. Zwaigenbaum, J. Brian, W. Roberts, P. Szatmari, V. Rombough, and C. McDermott. A prospective case series of high-risk infants who developed autism. *J. Autism Dev. Disord.*, 37:12–24, 2007.

[4] S. Chen, G. Gordon, and R. Murphy. A novel graphical model approach to segmenting cell images. *Proc. IEEE Symp. Comput. Intell. Bioinform. Comput. Biol. (CIBCB)*, 6:1–8, 2006.

[5] I. Cleary, P. Mulabagal, S. Reinhard, N. Yade, C. Murdoch, M. Thornhill, A. Lazzell, C. Monteagudo, D. Thomas, and S. Saville. Pseudohyphal regulation by the transcription factor rfg1p in candida albicans. *Eukaryotic Cell*, 9:1363–1373, 2010.

[6] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, San Diego, USA, 2005.

[7] G. Dawson. Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder. *Dev. Psychopathol.*, 20(03):775–803, 2008.

[8] S. Di Zenzo. A note on the gradient of multi-image. *Computer Vision, Graphics, and Image Processing*, 33(1):116–125, 1986.

[9] K. Downing, K. Concannon, V. Vogel-Farley, C.A. Nelson, and T Tager-Flusberg. Can the AOSI at nine months discriminate between infants at high or low risk for ASD? In *IMFAR*, San Diego, USA, 2011.

[10] R. Duda and P. Hart. Use of the hough transform to detect lines and curves in pictures. *Communications of the ACM*, 1972.

[11] M. Everingham, J. Sivic, and A. Zisserman. "Hello! My name is... Buffy" - Automatic naming of characters in TV video. In *BMVC*, Edinburgh, UK, 2006.

[12] M. Freeth, T. Foulsham, and P. Chapman. The influence of visual saliency on fixation patterns in individuals with spectrum disorders. *Neuropsychologia*, 49:156–160, 2010.

[13] M. Goodwin, S. Intille, F. Albinali, and W. Velicer. Automated detection of stereotypical motor movements. *J. Autism Dev. Disord.*, 41(6):770–782, 2011.

[14] P. Guan. Blood cell image segmentation based on the hough transform and fuzzy curve tracing. In *IEEE Conference on Machine Learning and Cybernetics*, volume 4, 2011.

[15] B. Harrison, J. Hashemi, G. Sapiro, and J. Berman. Morph than meets the eye: towards understand the mechanism of aneuploid formation in fluconazole-treated cells. In *American Society for Microbiology*, 2012.

[16] B. Harrison, J. Hashemi, G. Sapiro, and J. Berman. A tetraploid intermediate precedes aneuploid formation in fluconazole-exposed yeasts. In review, 2013.

[17] J. Hashemi, B. Harrison, J. Berman, and G. Sapiro. Automatic morphological and intracellular analysis of yeast cells. In review., 2013.

[18] J. Hashemi, T. Spina, M. Tepper, A. Esler, V. Morellas, N. Papanikolopoulos, and G. Sapiro. Automatically aiding the assessment of atypical motor patterns for autism spectrum disorder diagnosis. In *AACAP*, 2012.

[19] J. Hashemi, T. Spina, M. Tepper, A. Esler, V. Morellas, N. Papanikolopoulos, and G. Sapiro. A computer vision approach for the assessment of autism-related behavioral markers. In *ICDL-EpiRob*, San Diego, USA, 2012.

[20] J. Hashemi, T. Spina, M. Tepper, A. Esler, V. Morellas, N. Papanikolopoulos, and G. Sapiro. Computer vision tools for the non-invasive assessment of autism-related behavioral markers. In review, 2013.

[21] K. Ishida, J. Cola, F. Rodrigues, S. Cammerer, J. Urbina, I. Gilbert, W. de Souza, and S. Rozental. Synthetic arylquinuclidine derivatives exhibit antifungal activity against candida albicas, candida tropicalis and candida parapsilopsis. *Annals of Clinical Microbiology and Antimicrobials*, 10(3), 2011.

[22] R. Jafri and H. Arabnia. A survey of face recognition techniques. *Journal of Information Processing Systems*, 5(2):41–69, 2009.

[23] W. Jones, K. Carr, and A. Klin. Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder. *Arch. Gen. Psyschiatry*, 65:946–954, 2008.

[24] M. Kabir, M. Hussain, and Z. Ahmad. Candida albicans: a model organism for studying fungal pathogens. *ISRN Microbiology*, 2012.

[25] Z. Kalal, K. Mikolajczyk, and J. Matas. Face-TLD: Tracking-learning-detection applied to faces. In *ICIP*, Hong Kong, China, 2010.

[26] M. Kass, A. Witkin, and D. Terzopoulos. Snakes- active contour models. *International Journal of Computer Vision*, 1:321–331, 1987.

[27] F. Klauschen, M. Ishii, H. Qi, M. Bajenoff, J. Egen, R. Germain, and M. Meier-Schellersheim. quantifying cellular interaction dynamics in 3d fluorescence microscopy data. *Nature Protocols*, 4:1305–1312, 2009.

[28] A. Klin, W. Jones, R. Schultz, F. Volkmar, and D. Cohen. Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Arch. Gen. Psychiatry*, 59:809–816, 2002.

[29] A. Kuijper and B. Heise. An automatic cell segmentation method for differential interference contrast microscopy. In *ICPR 2008*, 2008.

[30] R. Landry and S. Bryson. Impaired disengagement of attention in young children with autism. *J. Child Psychol. Psychiatry*, 45(6):1115–22, 2004.

[31] M. Malcui and F. Preteux. A robust model-based approach for 3d head tracking in video sequences. In *IEEE Face and Gesture Recognition*, 2000.

[32] M. Marin-Jimenez, A. Zisserman, and V. Ferrari. "Here's looking at you, kid." Detecting people looking at each other in videos. In *BMVC*, Dundee, UK, 2011.

[33] M. Martin. The use of fluconazole and itraconazole in the treatment of candida albicans infections: a review. *Journal of Antimicrobial Chemotherapy*, 44:429–437, 1999.

[34] C. Michels. *Genetic Techniques for Biological Research: A case study approach*, chapter Saccharomyces cerevisiae as a genetic model organism. John Wiley and Sons, Ltd, 2002.

[35] E Murphy-Chutorian and M Trived. Head pose estimation in computer vision: a survey, 2008.

[36] F. Nazneen, F. Boujarwah, A. Sadler, Shone, G. Abowd, and R. Arriaga. Understanding the challenges and opportunities for richer descriptions of stereotypical behaviors of children with ASD: a concept exploration and validation. In *ASSETS*, Orlando, USA, 2010.

[37] N. Otsu. A threshold selection method from gray level histograms. *IEEE Trans. Systems, Man and Cybernetics*, 9:62–66, 1979.

[38] T. Peng. Detect circles with various radii in grayscale image via hough transform, 2010.

[39] D Prasher. Primary structure of the aequorea victoria green-fluorescent protein. *Gene*, 111(2):229–233, 1992.

[40] P. Rodier. Converging evidence for brain stem injury in autism. *Dev. Psychopathol.*, 14(03):537–557, 2002.

[41] A. Rosenfeld and J. Pfaltz. Sequential operations in digital picture processing. *Journal of the Association for Computing Machinery*, 13:471–494, 1966.

[42] P. Shattuck, M. Durkin, M. Maenner, C. Newschaffer, D. Mandell, L. Wiggins, L. Lee, C. Rice, E. Giarelli, R. Kirby, J. Baio, J. Pinto-Martin, and C. Cuniff. Timing of identification among children with an autism spectrum disorder: findings

from a population-based surveillance study. *J. Am. Acad. Child Adolesc. Psychiatry*, 48(5):474–483, 2009.

[43] M. Tepper and G. Sapiro. Decoupled coarse-to-fine matching and nonlinear regularization for efficient motion estimation. In *ICIP*, Orlando, USA, 2012.

[44] M. Vatahska, T. Bennewitz and S. Behnke. Feature-based head pose estimation from images. In *Humanoids*, 2007.

[45] V. Veses and NA. Gow. Pseudohypha budding patterns of candida albicans. *Med Mycol*, 47:267–275, 2009.

[46] A. Weissenfeld. Robust rigid head motion estimation based on differential evolution. In *IEEE Multimedia and Expo*, 2006.

[47] A. Yilmaz, O. Javed, and M. Shah. Object tracking: a survey. *ACM Comput. Surv.*, 38(4), 2006.

[48] J. Young, J. Locke, A. Altinok, N. Rosenfeld, T. Bacarian, P. Swain, E. Mjolsness, and M. Elowitz. Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy. *Nature Protocols*, 7:80–88, 2011.

[49] L. Zwaigenbaum, S. Bryson, T. Rogers, W. Roberts, J. Brian, and P. Szatmari. Behavioral manifestations of autism in the first year of life. *Int. J. Dev. Neurosci.*, 23(2-3):143–152, 2005.