Internship Report

Improving estimation of grasping polygons using gaze information

At Okayama University

Mr. Tanakan Pramot
Student ID 5810502334

Academic Year 2018
Department of Computer Engineering, Faculty of Engineering
Kasetsart University

Abstract

The purpose of the report is to explain the details of writer's internship program at Okayama University, Japan in 2018. This internship is part of the bachelor's degree of computer engineering program.

The topic of my project is "Improving estimation of grasping polygons using gaze information". I learned new methods to work on research since I did not have any experience and this is my first research project.

In this internship program, I gave 2 presentations, which are mid-term and final presentations to show the progress and results of my work to members of Monden laboratory which is my laboratory, and Yamauchi laboratory.

Acknowledgement

The internship report is the overall result of the Internship program at Monden laboratory, Okayama University, Japan. The internship is part of the bachelor’s degree of computer engineering program.

My internship project is improving estimation of grasping polygons using gaze information. The ultimate purpose of my project is to teach a robot how to grasp objects like a human does. For this purpose, we first need to define graspable parts of the objects. To define the graspable parts, I used human gaze information to improve estimation of grasping polygons of an object.

I would like to thank the people who made this internship possible. Asst. Prof. Zeynep Yucel, Prof. Akito Monden, Asst. Prof. Pattara Leelaprute, Assoc. Prof. Arnon Rungsawang, Asst. Prof. Bundit Manaskasemsak provided valuable assistance and and the members of Monden laboratory supported my stay in Japan.

I hope this experiences and the knowledge that I gained from the internship will improve my future work and be guidance to any young students further.

Mr. Tanakan Pramot
Reporter

The last date of internship
31/7/2018

Contents

List of Figures

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Motivations and Importance

Recently, robots can replace humans in many specialized tasks, such as technical tasks in a factory. However, robots are still far from doing many tasks, which are very simple for humans. Therefore, humans need to train robots more by using information from themselves to improve robot's performance.

1.2 Objectives

1. Improve robot grasping performance by using human's gaze information.
2. Teach robot to grasp like a human. (Future work)

1.3 Scope of Work

This outilne of the study canbe summarized as analysis of a ground truth dataset, formulation of the problem in an analytical way, software development on a computer platform and proof of improvement in performance.

1.4 History and Detail of Company

Figure 1 Logo of Okayama University

Name: Okayama University

Description: The laboratoy is called ‘Software Measurement and Analytics Laboratory’ (a.k.a Monden Lab) and is part of the Graduate School of Natural Science and Technology, Okayama University. The main purpose of the laboratory is scientific research. Supervisor checks on research progress by having an appointment with researcher (or student) several times in a week.

Position: Research student

Supervisor: Prof. Akito Monden (Internship supervisor)
Asst. Prof. Zeynep Yucel (Project consultant)

Period: 1st June – 31st July 2018

1.5 Expected Benefits

1. Experience and knowledge about working as a researcher in computer science field.
2. Learning about studying abroad in master’s degree and abroad life.
3. Get connections with foreign friends.

Chapter 2
Background Knowledge and Related Work

2.1 Background Knowledge

In daily life, humans interact with various objects. From an early age, humans develop motor and cognitive skills to manage their interaction with objects. These possible actions that objects can be employed for are called as "Affordance ".

Robots or computer programs cannot understand "Affordance " by default. Therefore, humans need to teach them by transferring them their own knowledge. For this purpose, we use the "Attention" of humans, which they direct twords objects during their interaction with them. "Attention" of humans can be determined from their movement, gaze, or brain activities.

Beside knowing about "Affordance" and "Attention", we need to carry out software coding to implement the solution.

2.2 Related Work

We use the freely available "Learning to grasp" data set of Cornell University, which contains 1034 images of a variety of graspable objects from various orientations.

Figure 2 Examples of Cornell University data set

For each image, several grasping polygons are annotated by human coders. Grasping polygons are not always imagined for human grasping, but they certainly do not exclude that either.

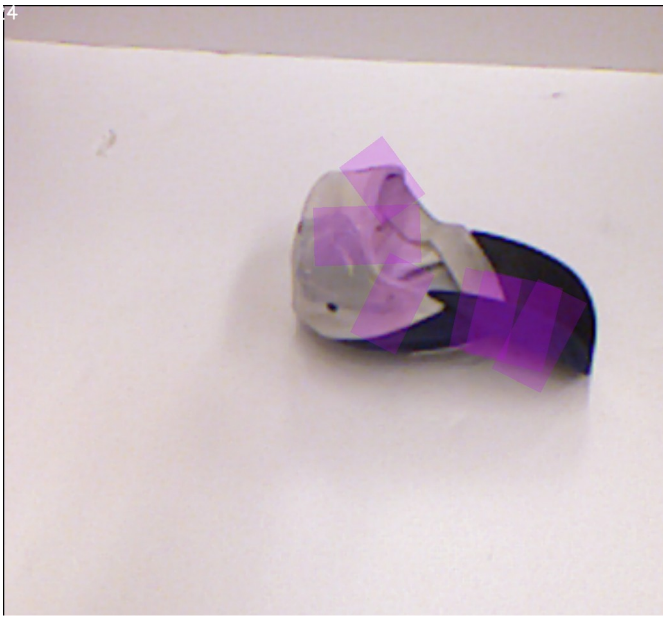

Figure 3 Example of grasping polygons

Chapter 3
Methodology

3.1 Experiment

3.1.1 Experiment video

A random subset of 432 images are selected, cropped, and resized to 1920 x 1080 resolution, to reduce clutter and provide a clear view. 12 video clips are built, displaying a series of 36 images each (3 seconds for each image followed by a 2 seconds reset period).

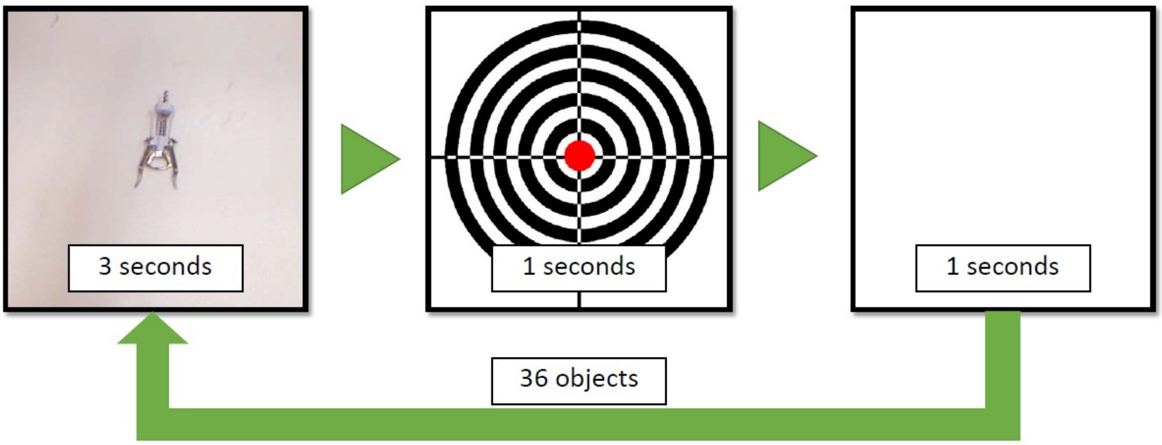

Figure 4 Order of pictures in video clip

3.1.2 Subjects

Two subjects are instructed to image grasping the objects (i.e. motivated) and the other two carry out a free viewing task (i.e. unmotivated). Which mean motivated is watching for grasping and unmotivated is watching for curiosity.

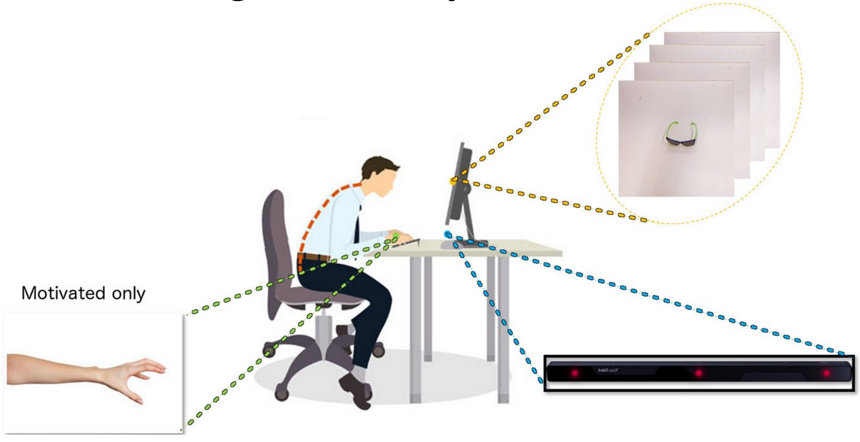

Figure 5 Experiment scenario

3.2 Gaze information

3.2.1 Gaze data

Gaze information is collected by an infrared sensor at a frequency of 70 Hz, and the gaze coordinates are logged together with corresponding time stamps.

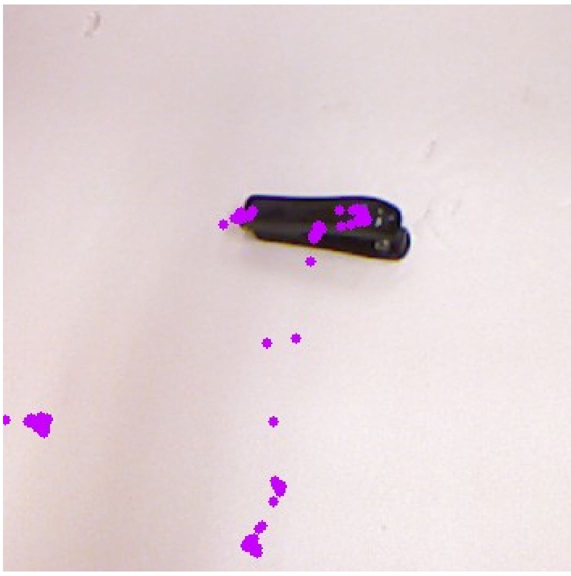

Figure 6 Gaze point on image

After we got gaze data we want to:

- Get a continuous map (i.e. heat map) from discrete coordinates
- Identify fixations (i.e. connected regions in the heat map)
- Segment out the fixations (i.e. blobs) for modeling gaze

3.2.2 Heat map

We build a continuous heat map from discrete coordinates by applying a 2D Gaussian kernel at each gaze coordinate (g_x, g_y).

$$k_g(x,y)=\exp\left(\frac{-(x-g_x)^2+(y-g_y)^2}{2\sigma^2}\right)$$

Saliency of the screen coordinates (x,y) are assumed to be independent, i.e. co-variance matrix is symmetric.

$$\sigma_{xy}=\sigma_{yx}=0$$

Moreover, co-variance matrix is assumed to be isotropic (i.e. spherical).

$\sigma_{xx} = \sigma_{yy} = \sigma$

After we compute $k_g(x,y)$ with support of 101x101 px with a proper σ (chosen using 3σ rule), we will get heat map from discrete gaze coordinates.

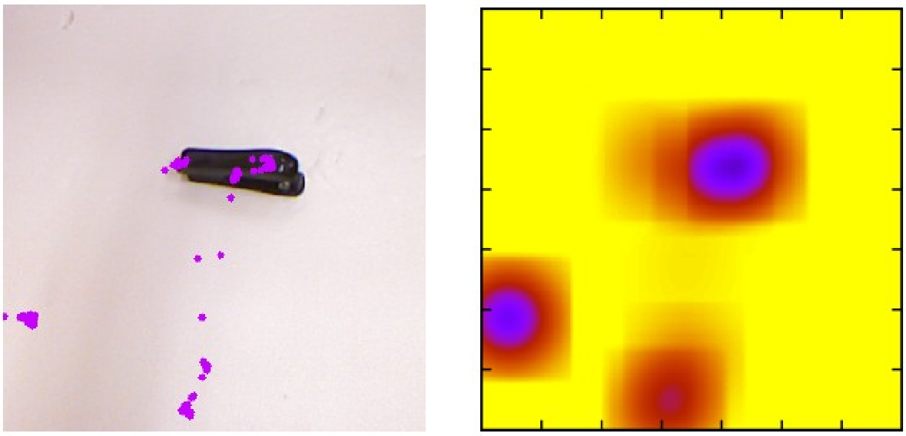

Figure 7 Discrete coordinates and Heat map

3.2.3 Segmenting fixations

After we got heat map, we will segment out the gaze fixation (i.e. blobs) by binarize the heat map using Otsu's adaptive thresholding.

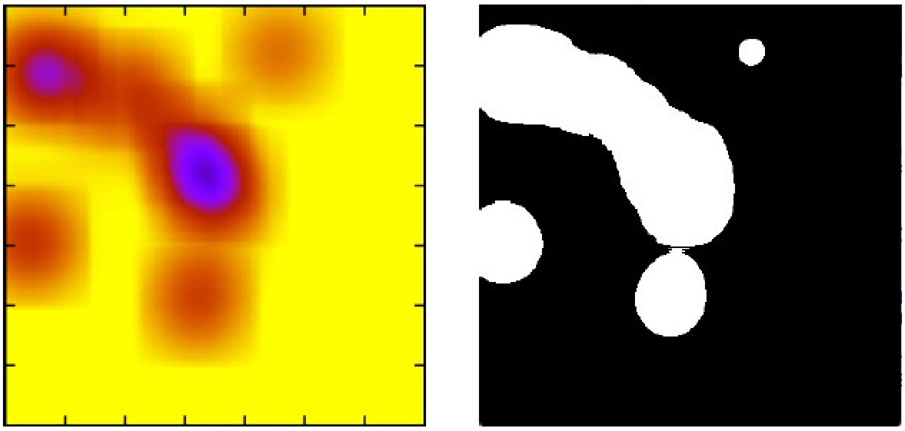

Figure 8 Heat map and Segmented gaze fixations

Now, we have gaze fixations that we can working on it instead of using all discrete coordinates.

3.3 Foreground estimation

3.3.1 Background subtraction

First of all, we will extract object from background image to get the foreground image in three channels. This process is called as "Background subtraction". We get RGB foreground image after doing this process.

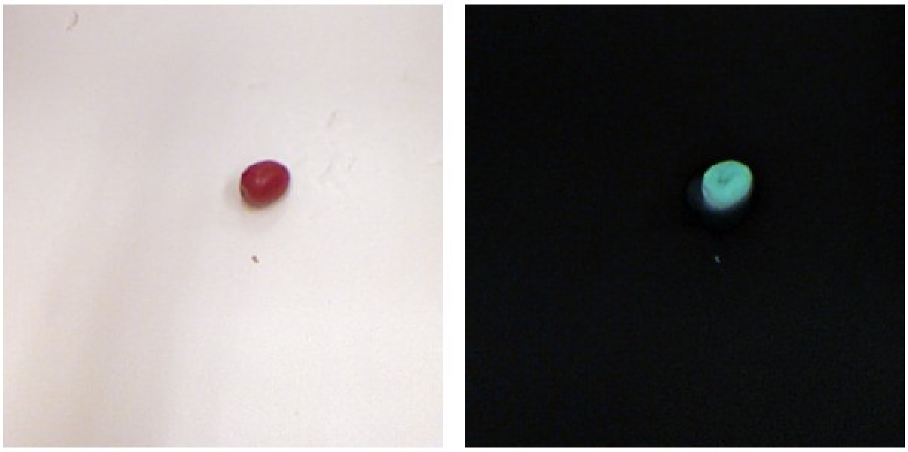

Figure 9 Before and after doing background subtraction

3.3.2 Grayscale conversion

We convert RGB image into grayscale image by taking a linear combination of its channels.

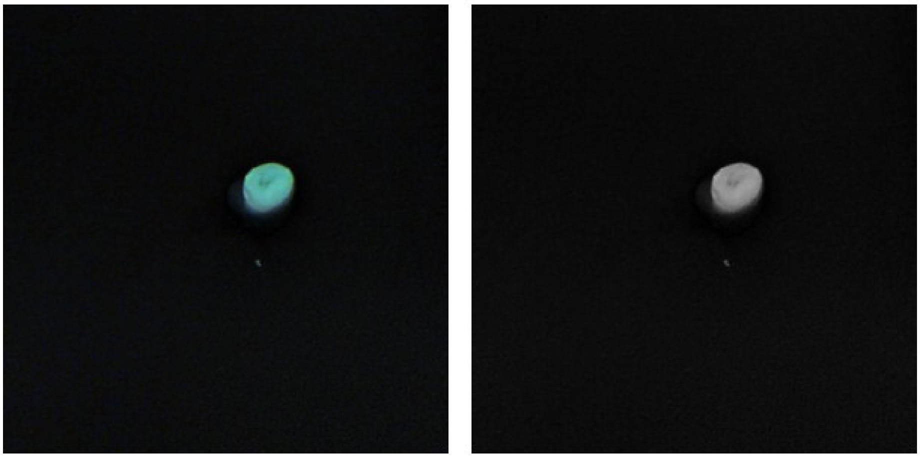

Figure 10 Before and after doing grayscale conversion

3.3.3 Hard-thresholding

Hard-thresholding of grayscale image does not always yield a nice binarization. We need to apply some special method instead of Hard-thresholding.

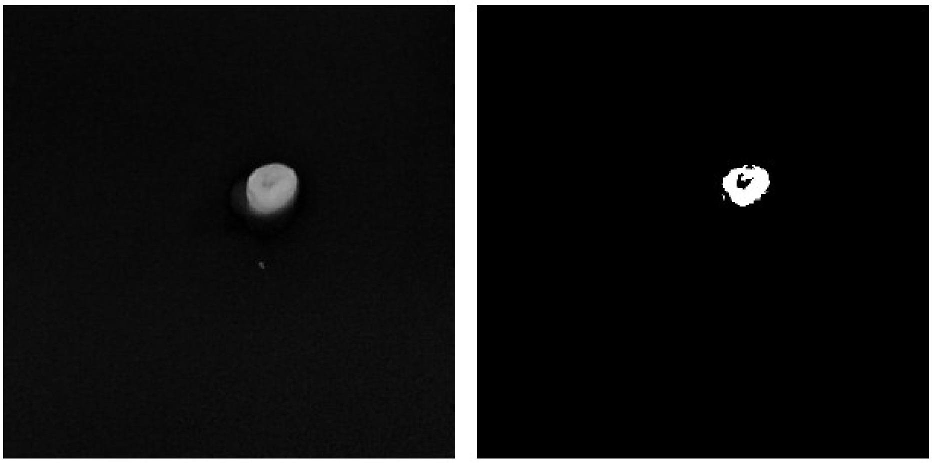

Figure 11 Before and after doing Hard-thresholding

3.3.4 Otsu's method

We apply Otsu's adaptive filter to get a better binary image.

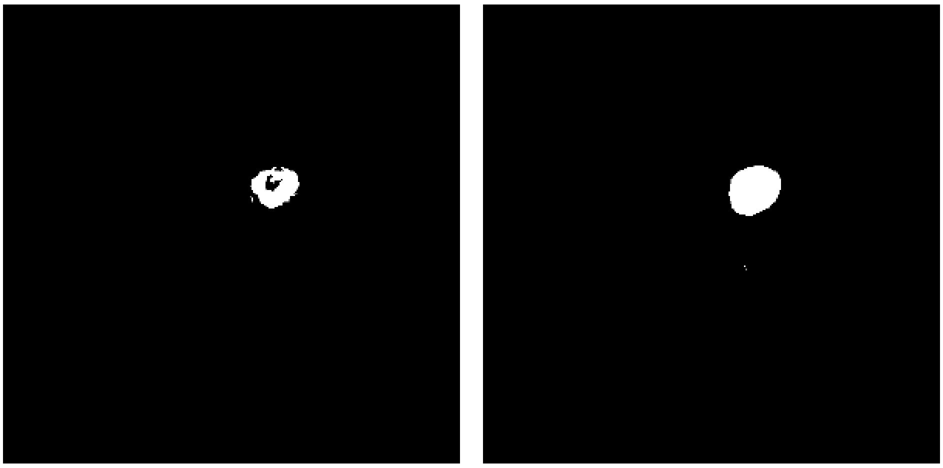

Figure 12 Before and after applying Otsu's adaptive filter

Now, we have segmented out the area of the object that we can working on.

3.4 Grasping polygons

We have many objects in dataset, and all them are graspable so there is at least one grasping polygon for each.

Despite we call them polygons, but they are actually rectangular regions. Two kinds of polygons are annotated, positive polygons (Conventional grasp) and negative polygons (Unconventional grasp). In this research, we focus on positive polygons.

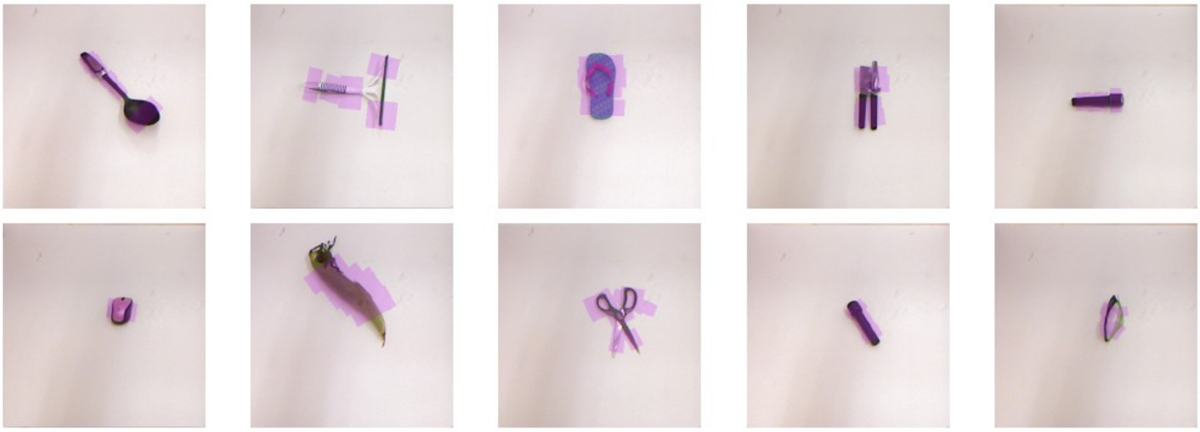

Figure 13 Grasping polygons for various objects

3.5 Descriptors

We need to describe gaze, object and polygon properties to represent their morphology and orientation. Descriptors are crucial in exploring for the relations and correlations between grasping intuitions, and gaze behavior and object properties.

3.5.1 Object descriptors

These descriptors help us understand about the shape and orientation of the object.

- Object's size S_o

S_o is the first moment of the binary foreground image.

$$S_o = \sum_x \sum_y F_B(x, y)$$

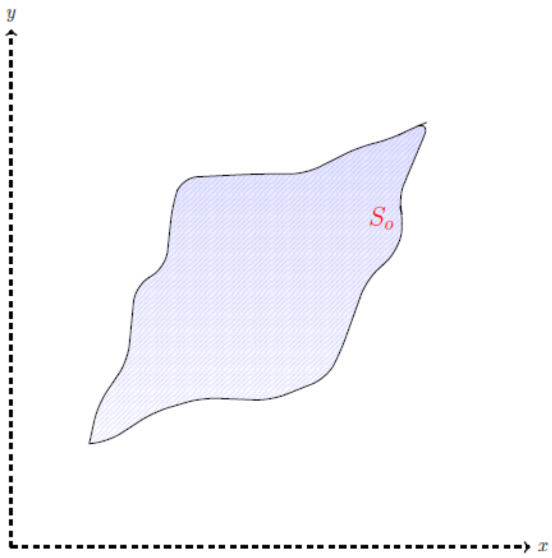

Figure 14 Example object in xy coordinates

- Orientation θ_o

We use an elliptic model to approximate θ_o. We use Eigen decomposition of binary image (F_B) to calibrate the elliptic model.

$$F_B v_{p,s} = \lambda_{p,s} v_{p,s}$$

Principal and secondary axes $r_{p,s}$ are aligned with eigen vectors of F_B , $v_{p,s}$ where $\lambda_p > \lambda_s$. θ_o is the slope of the principal axis of the ellipse,

$$\theta_o = \arctan 2(\hat{v}_{pv}, \hat{v}_{px})$$ Where $\hat{v}_{px,y}$ are components of the unit vector $\hat{v}_p$

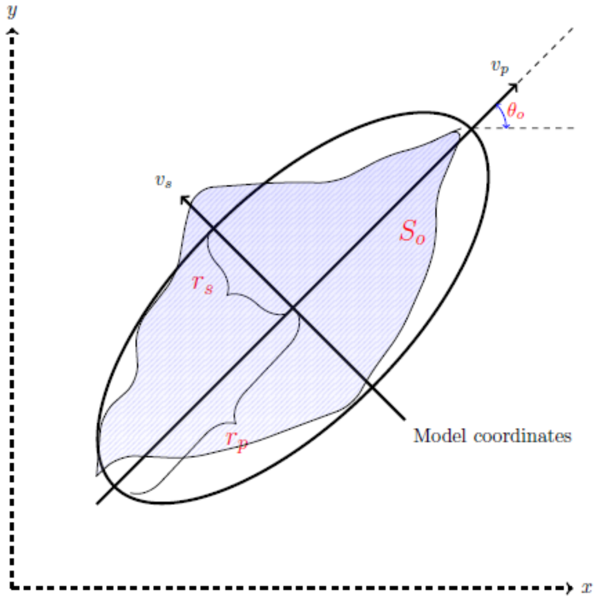

Figure 15 Elliptic model

- Inertia ratio R_o

R_o is the ratio of secondary axis to the principal axis, i.e. intuitively smaller eigen value to the larger one,

$$R_o = \frac{\lambda_s}{\lambda_p}$$

But eigen value decomposition results in overestimation of $\vee \acute{r}_{p,s} \vee$

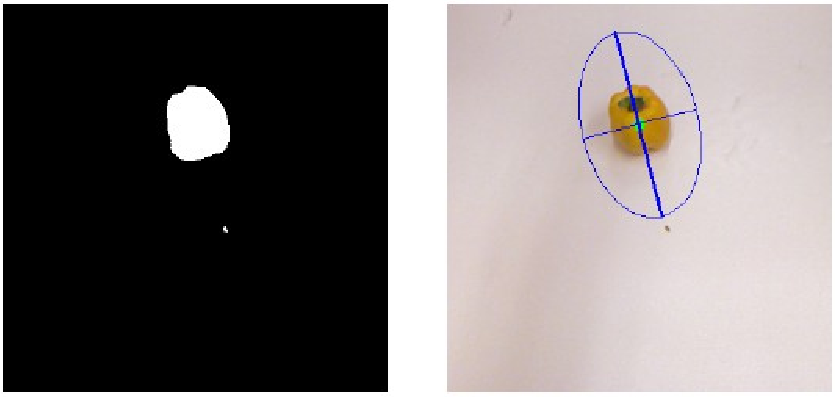

Figure 16 Overestimation from eigen value decomposition

Therefore, we rotate the image by $-\theta_o$ and compute the horizontal and vertical projections of this rotated image,

$$\tilde{F}_B^x = \sum_y \tilde{F}_B(x,y) \quad \tilde{F}_B^y = \sum_x \tilde{F}_B(x,y)$$

This can be used to approximate $\| r_{p,s} \|$ through,

$$\vee \tilde{r}_p \vee = \sum sgn\left(\tilde{F}_B^x\right) \vee \tilde{r}_s \vee = \sum sgn\left(\tilde{F}_B^y\right)$$

Then, $R_o = \tilde{r}_s / \tilde{r}_p$ and limited to the range (0,1].

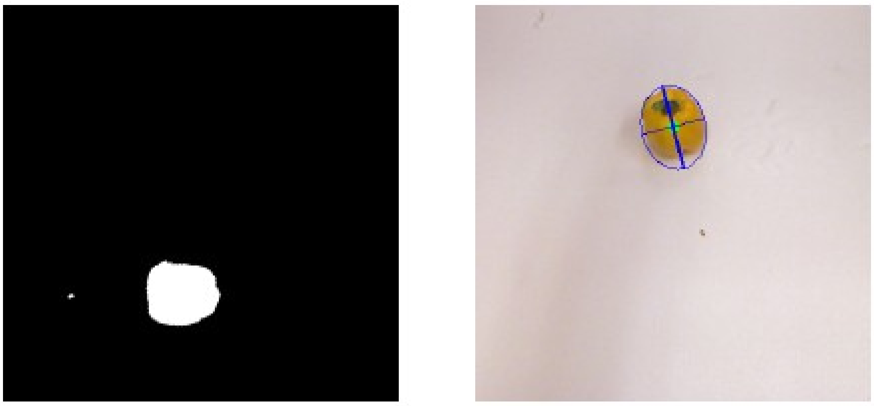

Figure 17 Rotated binary image and its ellipse

3.5.2 Gaze descriptors

Gaze fixations are represented with blobs on the heat map which can be treated in the same way as the object on binary image. (Elliptic models) But this time we have two kinds of descriptors.

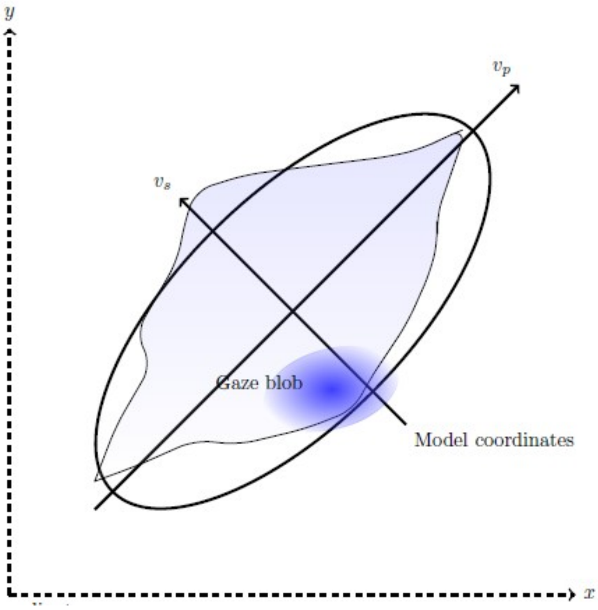

Figure 18 Gaze blob

3.5.2.1 Morphological

- Blob's size S_g

S_g is the first moment of each blob on the fixation map. (Similar to S_o.)

- Orientation θ_g

θ_g is defined according to the elliptic model of the blob.

$$\theta_g = \arctan 2\left(\hat{v}_{gp} / \hat{v}_{gs}\right)$$

- Inertia ratio R_g

R_g is defined according to the elliptic model of the blob.

$$R_g = \tilde{r}_{gs} / \tilde{r}_{gp}$$

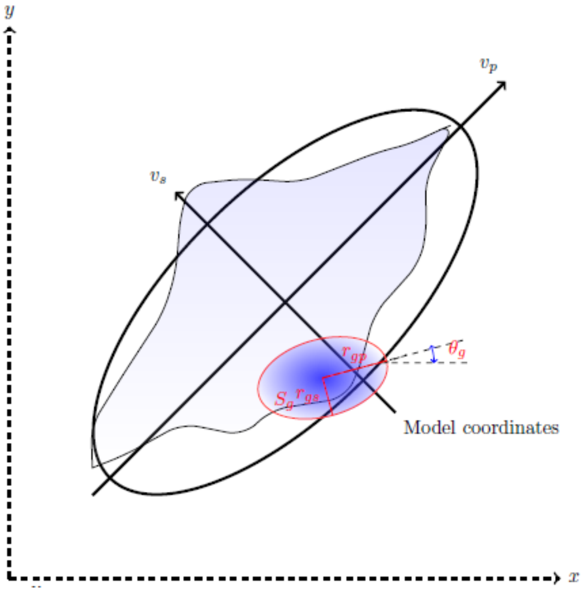

Figure 19 Gaze's morphological descriptor

3.5.2.2 Positional

- Radial distance δ_g

The radial distance is defined in a normalize manner as,

$$\delta_g = \frac{\vee C_o - C_g \vee}{S_o}$$

Where $C_{o,g}$ are centroids of the object and gaze blob. This normalization helps to make δ_g independent of the area available.

- Angular position ω_g

Angular position or ω_g is defined as,

$$\omega_g = \arccos\left(\frac{\vec{r}_p\left(\vec{C}_o - \vec{C}_g\right)}{\vee \vec{r}_p \vee \vee \vec{C}_o - \vec{C}_g \vee}\right)$$

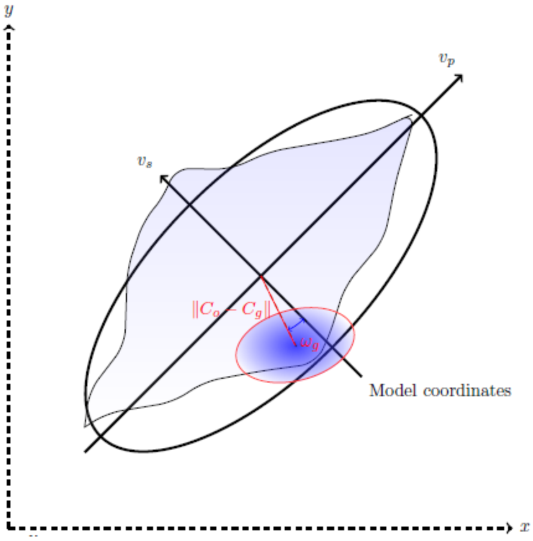

Figure 20 Gaze's positional descriptor

3.5.3 Polygon descriptors

3.5.3.1 Morphological

Polygons have analogous morphological descriptors.

- Polygon's size S_p

$$S_p = wh$$

- Orientation θ_p *(We are using $\Delta\theta$ instead of θ later)

$$\theta_p = \arctan 2\left(\vec{w}/\vec{h}\right)$$

- Inertia ratio R_p

$$R_p = h/w$$

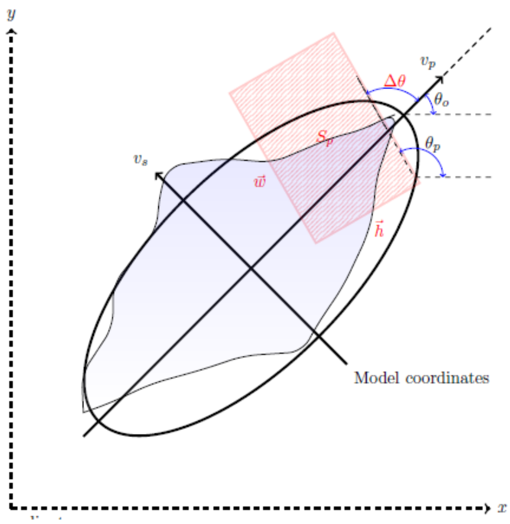

Figure 21 Polygon's morphological descriptor

3.5.3.2 Positional

The positional descriptors are analogous as well.

- Radial distance δp

$$\delta_p = \frac{\vee C_p - C_o \vee}{S_o}$$

- Angular position ω_p

$$\omega_p = \arccos\left(\frac{\vec{r}_p\left(\vec{C}_p - \vec{C}_o\right)}{\vee \vec{r}_p \vee \vee \vec{C}_p - \vec{C}_o \vee}\right)$$

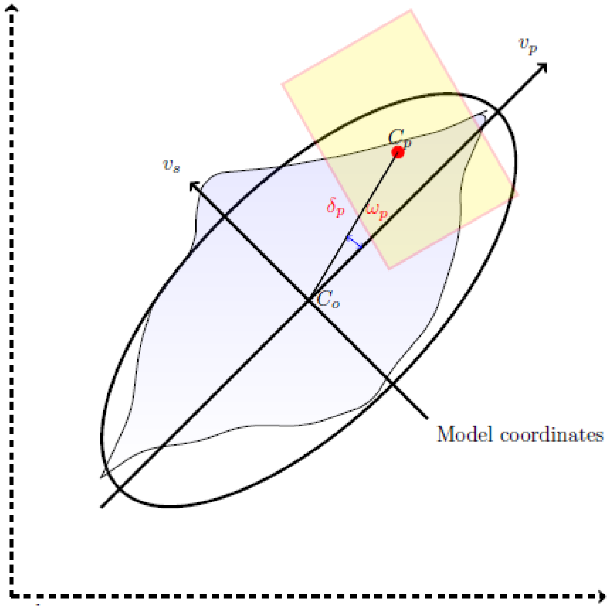

Figure 22 Polygon's positional descriptor

3.6 Empirical relations and models

After we have all of gaze, object, and polygon properties, we expect the following pairs to have a certain degree of correlation:

- S_o and $S_{p,g}$
- R_o and $R_{p,g}$
- θ_o and $\theta_{p,g}$ ($\Delta\theta$)
- R_o and δp,g
- R_o and $\omega_{p,g}$

3.6.1 Object size and polygon size: S_o and S_p

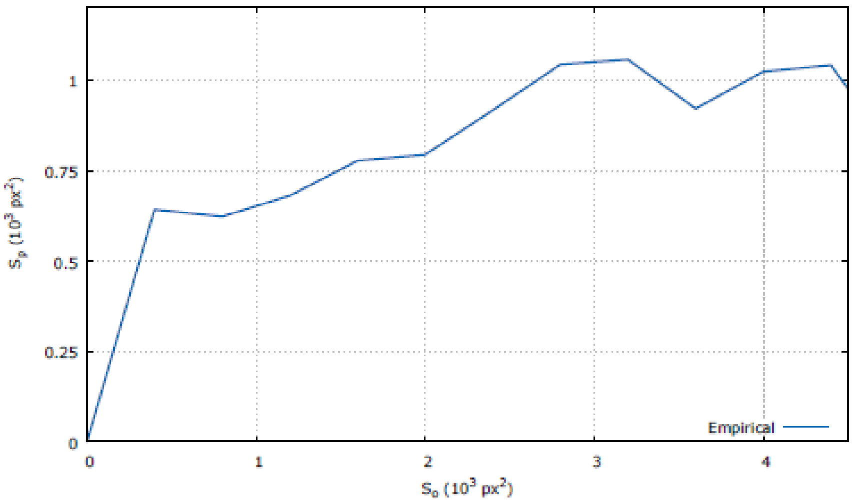

Figure 23 Relation of S_o and S_p

It is not surprising that grasping polygons of small objects are small and the size grows in a somewhat negative exponential relation with respect to the size of the object settling around a stable value for very large objects.

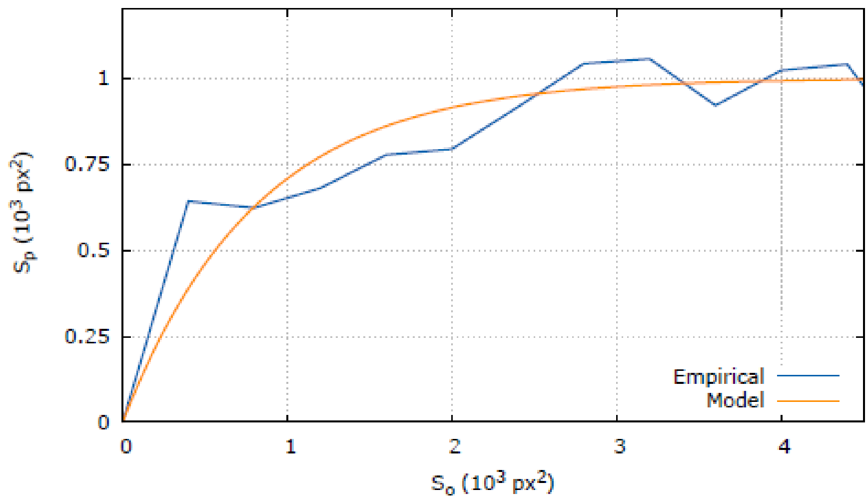

Figure 24 Exponential model of S_o and S_p

We use an exponential model:

$$f(S_p | C_s, \lambda) = C_s (1 - \exp(-\lambda S_o))$$

Where C_s is the scaling factor, λ determines the rate of decay and S_o stands for the size of the object.

3.6.2 Object inertia ratio and polygon inertia ratio: R_o and R_p

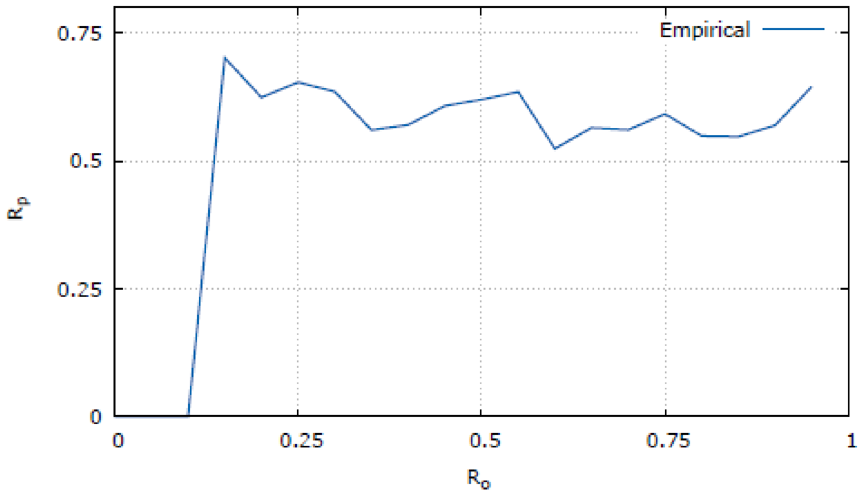

Figure 25 Relation of R_o and R_p

There is no apparent relation between R_o and R_p. However, R_p seems to be quite stable against R_o.

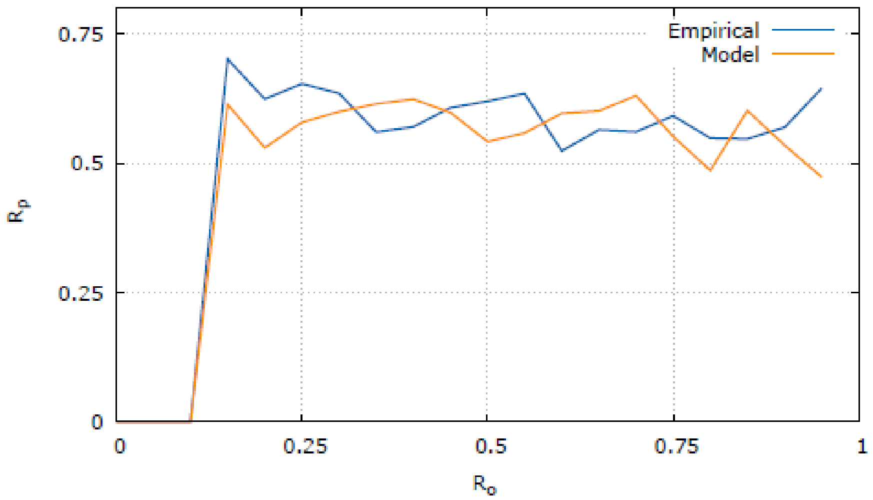

Figure 26 Normal distribution model of R_o and R_p

We assume $R_p \sim N(\mu_{rp}, \sigma_{rp})$, where μ_{rp} and σ_{rp} are computed from the empirical observations.

3.6.3 Object orientation and polygon orientation: θ_o and θ_p ($\Delta\theta$)

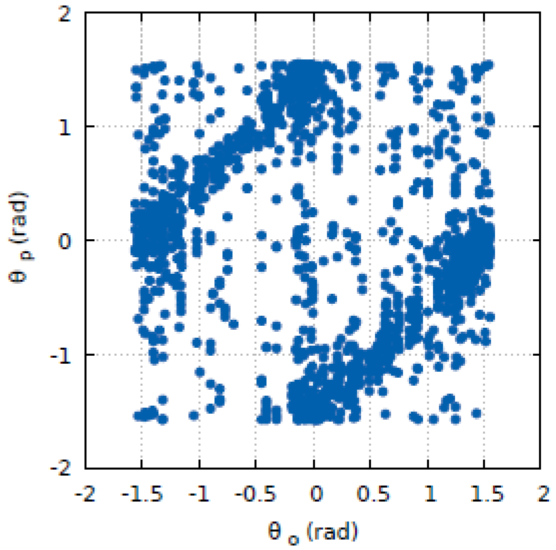

Figure 27 Relation of θ_o and θ_p

We see a clear pattern that there is a certain offset between θ_o and θ_p. Thus we decide to examine $\Delta\theta = \theta_o - \theta_p$.

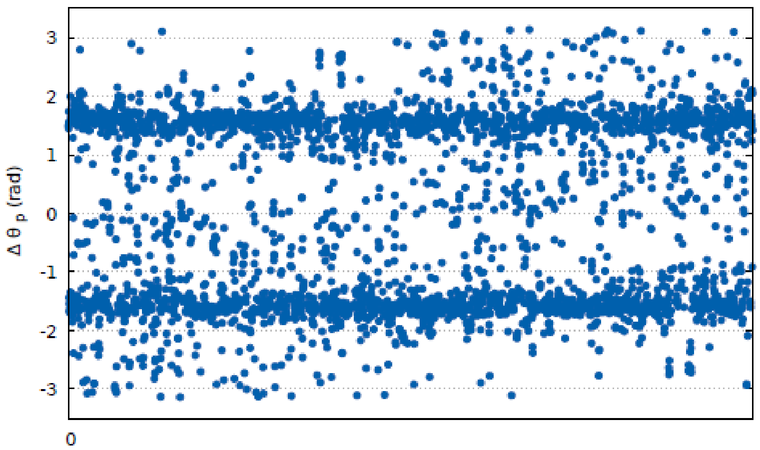

Figure 28 Orientation difference plot

$\Delta\theta$ suggests that principal axes of objects and polygons are often perpendicular to each other. We decided to compute the cumulative histogram and approximate it with Von Mises distribution.

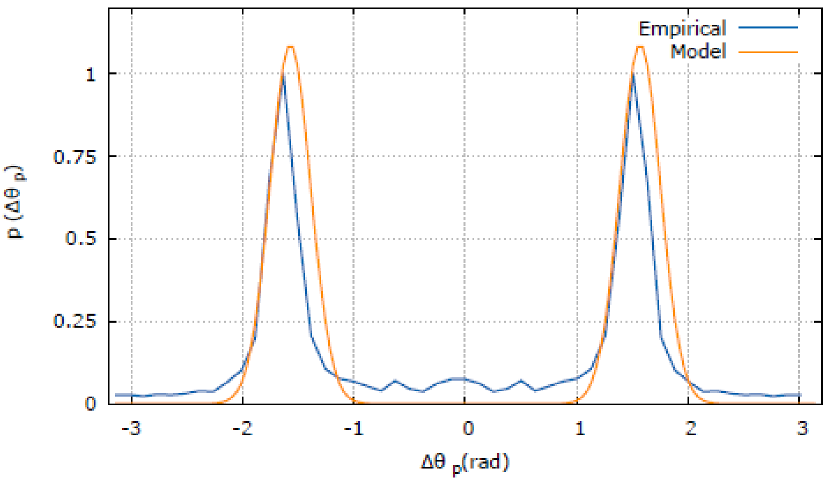

Figure 29 Von Mises distribution of Δθ

Von Mises distribution is the circular equivalent of the standard normal distribution,

$$f\left(\Delta\theta_p \vee \mu_{\Delta\theta_p}, \kappa_{\Delta\theta_p}\right) = \frac{\exp\left(\kappa_{\Delta\theta_p}\left(\Delta\theta_p - \mu_{\Delta\theta_p}\right)\right)}{2\pi I_0\left(\kappa_{\Delta\theta_p}\right)}$$

Where $\mu_{\Delta\theta p}$ and $\kappa_{\Delta\theta p}$ are analogous to the mean and standard deviations of the normal distribution and I_0 is the modified Bessel function of order 0.

3.6.4 Object inertia ratio and polygon normalized radial distance: R_o and δp

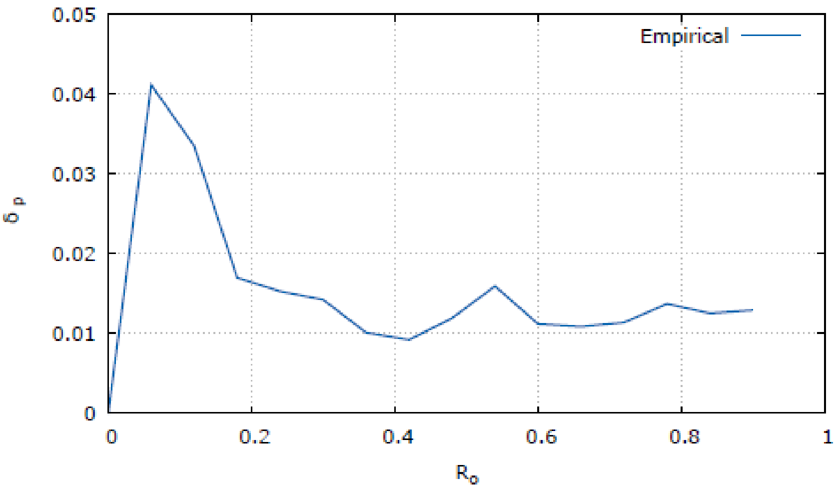

Figure 30 Relation of R_o and δp

This curve indicates that **δp** is larger for long objects and smaller for round objects. For modeling this relation, we assume that the projections of **δp** on r_p and r_s come from a normal distribution with a nonzero mean.

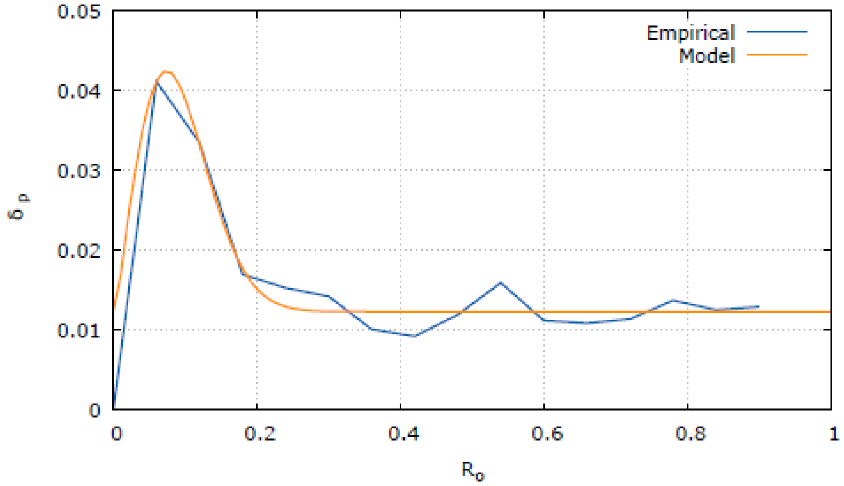

Figure 31 Rice distribution of R_o and δp

Then, δp respects a Rice distribution, which is the magnitude of a circular bivariate normal random variable with potentially non-zero mean.

$$f(\delta_p \vee \eta, \sigma) = C_r\left(\tau + \frac{\delta_p}{\sigma^2}\exp\left(\frac{-\left(\delta_p^2+\eta^2\right)}{2\sigma^2}\right) I_0\left(\frac{\delta\eta}{\sigma^2}\right)\right)$$

Where C_r is a scaling factor, τ is an offset value; and η and σ are the model parameters.

3.6.5 Object inertia ratio and polygon angular distance: R_o and ω_p

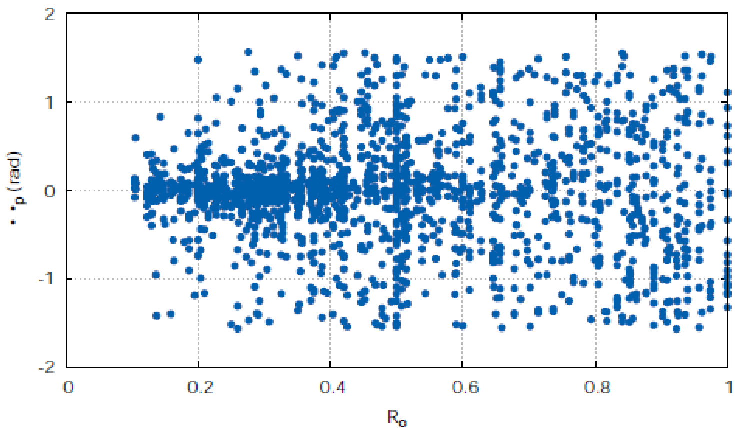

Figure 32 Relation of R_o and ω_p

Objects with small R_o are expected to have a more pronounced angular position than the ones with larger R_o. This expectation is supported by the empirical findings. A Von Mises distribution with an additional parameter for R_o would be the optimal. However, the calibration process of such a model is likely to suffer from the sparsity of the space.

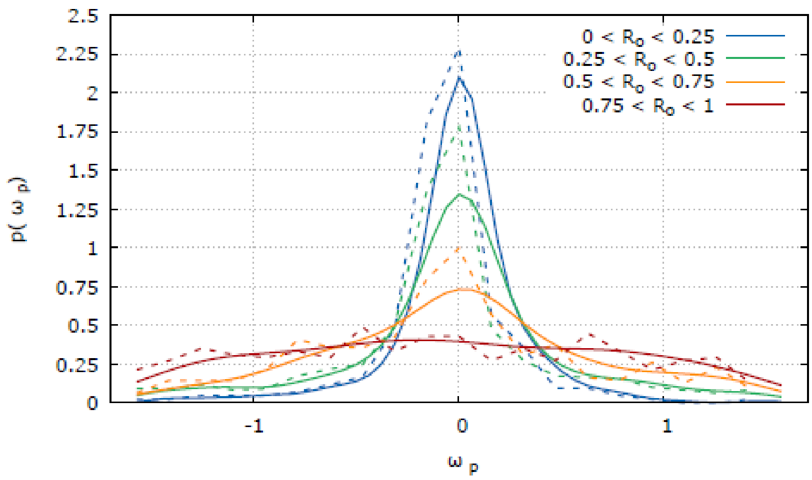

Figure 33 Von Mises distribution of R_o and ω_p

Therefore, we propose binning the distribution in four equal ranges of R_o and treating each of them with Von Mises distributions (of same parameters).

3.6.6 Object size and gaze size: S_o and S_g

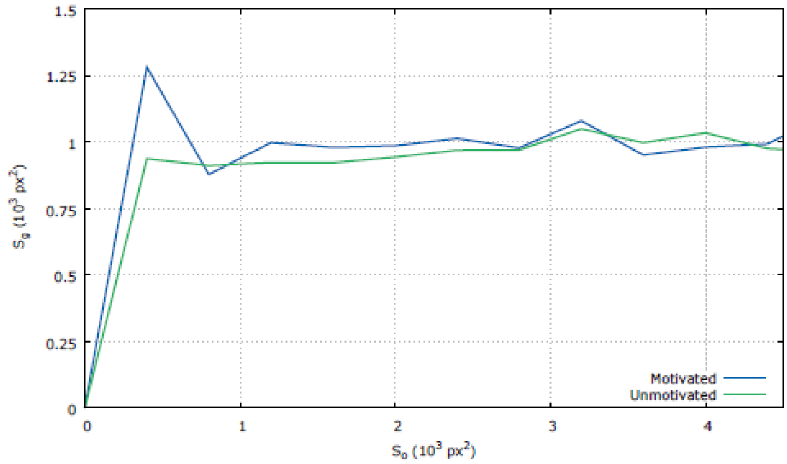

Figure 34 Relation of S_o and S_g

Motivated and unmotivated subjects present no significant difference. Moreover, S_g does not seem to depend on S_o to a large extend. Therefore, we use a standard normal distribution.

$S_g N\left(\mu_{sg}, \sigma_{sg}\right)$

3.6.7 Object inertia ratio and gaze inertia ration: R_o and R_g

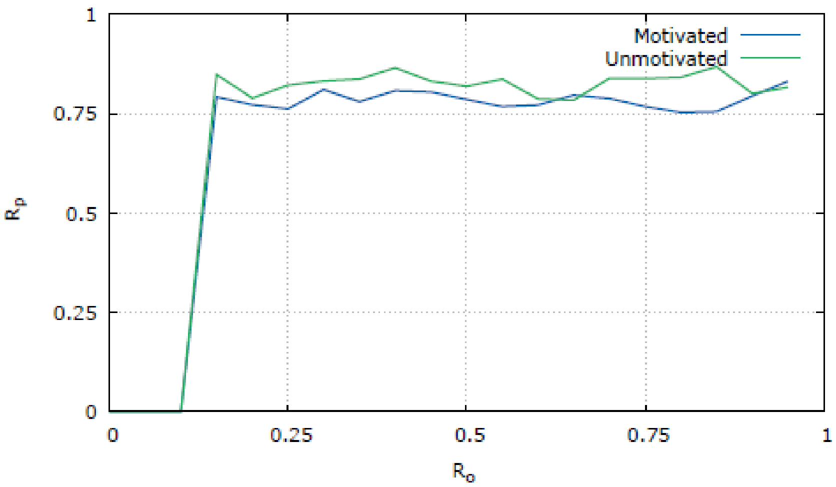

Figure 35 Relation of R_o and R_g

Motivated and unmotivated subjects present no significant difference. Moreover, R_g does not seem to depend on R_o. Therefore, we use a standard normal distribution.

$R_g N(\mu_{sg}, \sigma_{sg})$

3.6.8 Orientation difference: Δθ

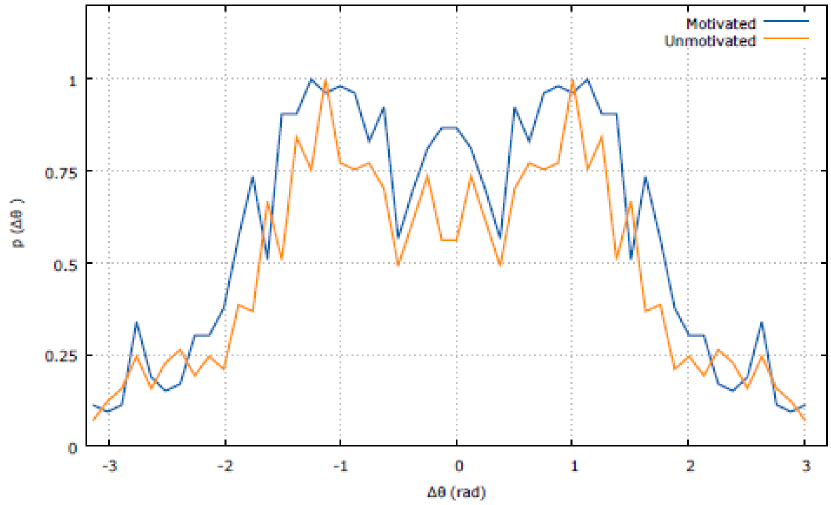

Figure 36 Relation of Δθ

Unmotivated subjects present slightly better correlation with polygon orientations. We use a Von Mises distribution model.

3.6.9 Object inertia ratio and gaze radial distance: R_o and δg

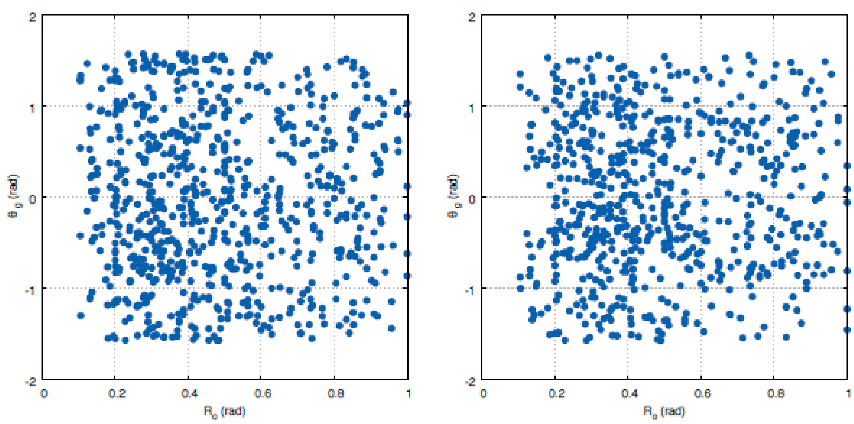

Figure 37 Relation of R_o and δg

Neither motivation nor R_o does not seem to have any effect on δg

3.7 Estimation of grasping polygons

3.7.1 Strategies

- We divide the set and apply 3-fold cross-validation.
- The training set is used to calibrate the models.
- For each image in the test set, one grasping polygon is generated.
- Since the grasping polygon is defined by its descriptors, we draw a random variable from the models of each descriptor.
- We use S_o and R_o of object for drawing random variables of S_p, R_p, θ_p, δp and ω_p.

3.7.2 Evaluation metric

We use the conventional intersection over union (IoU) metric for evaluating the performance. IoU is equivalent of Jaccard distance in the non-parametric form.

$$IoU = \frac{Mutual\,information}{Joint\,entropy}$$

In addition, we count the number of hits and misses. Misses are the instance where the estimation does not have an intersection with any annotated polygon.

Chapter 4
Results and Analysis

4.1 Results

4.1.1 Estimation performance using gaze information

	Hit	Miss	IoU
Gaze of motivated	775	89	0.23
Gaze of unmotivated	782	82	0.32

Table 1 Performance of gaze information

The gaze of unmotivated subjects yields better performance than motivated subjects in a consistent manner. Although this is a surprising finding, it was somewhat expected from the empirical distributions.

4.1.2 Estimation performance using heuristics

	Hit	Miss	IoU
Heuristic	688	176	0.24

Table 2 Performance of heuristic

Heuristics yield approximately same IoU as estimation from motivated gaze, but the miss rate is much higher. This means that when there is a hit, the IoU should be better than
motivated gaze.

On the other hand, it performs much worse than unmotivated gaze both in terms of IoU and hit rate.

4.1.3 Estimation performance of blending gaze and heuristics

	Hit	Miss	IoU
Gaze of motivated	775	89	0.23
Gaze of unmotivated	782	82	0.32
Gaze of motivated + heuristics	864	0	0.23
Gaze of unmotivated + heuristics	864	0	0.32

Table 3 Performance of blending gaze and heuristics

We estimate grasping polygons using gaze information. If the centroid of the estimated polygon is not over the foreground, we replace this estimation with heuristics.

4.2 Analysis

4.2.1 Using only gaze information or heuristics

As we can see on the Table 1, we expect that gaze of motivated will have better performance than gaze of unmotivated. However, we got result that unmotivated have better performance than motivated. The reason can be due to the active exploration of motivated subjects for planning actions because when subjects imagine that they are grasping object, they need to look around object to find efficient graspable parts of objects. This make gaze of motivated are more outside objects and give less performance than unmotivated.

In the Table 2, we can see that we have many misses more than using gaze information. The reason is when coder annotated the grasping polygons, some object has complicate shape. It makes some centroid of grasping polygons are outside of the objects.

4.2.2 Using both gaze information and heuristics

After we combine using gaze information with heuristic in the Table 3, we get more IoU than using only one of them. This is because when we use gaze information, it helps to define clearly grasping part of the object. The reason is subject will looking mostly in the foreground and same area in an object, so this will give better random variable when we generate the polygon.

Chapter 5
Conclusion

5.1 Work conclusion

The conclusion of this study is that gaze information can improve grasping polygons and yield better evaluation performance. This means that we can use our work as a baseline to teach robot to grasp like a human do as a future work.

5.2 Expectation

I expect that my internship project can be useful resource for researchers working in robotics or artificial intelligence fields.

5.3 Benefits

- Benefits to myself

I gain many experience from this internship such as experience in working as a researcher, experience in studying abroad, and experience in having foreign friends. I also get a chance to learning new things and have a chance to consider about master's degree in computer science, Okayama university.

- Benefits to company

They have connection with me, and I think we have more strong relationship between Kasetsart University and Okayama University. They also learn some Thai culture in return as they teach me about Japanese culture.

- Benefits to university

Similar to a company, university gets stronger relationship with Okayama University and hopefully the two universities will have more opportunities to have joint research projects in the future.

5.4 Swot Analysis

- Strong point

I'm very good in communicating in English language. I can easily communicate to people, who can understand English. I'm friendly man so I met so many people from many nations. While I am working, I'm working hard. While I am relaxing, I'm playing hard.

- Weak point

I can’t communicate with Japanese language, so I missed many opportunities to explore Japan by myself. I’m stubborn person sometimes. When I stuck in some work, I will try my best to solve it even it takes a long by not asking someone.

- Opportunity

I have many opportunities to learn new things. I also have opportunity to meet many people from many nations like French, Chinese, Vietnamese, Turkish, and Japanese.

- Obstacle

Language barrier between me and people who can’t talk English. Sometimes I am lazy, so I miss opportunity to travel around Japan because I decide to sleep in my dormitory than go out.

5.5 Impressive experience

All people I met are very nice. They are very kind and friendly. People in my laboratory are very kind, friendly, and funny. Especially, French guys. Not all people in my laboratory talk to me because they don’t want to speak English but if I talk to them, they will try to answer me kindly and slowly to make me understand. My two sensei, Zeynep-sensei, and Monden-sensei are very kind too. Monden-sensei is my internship supervisor, he will take care of me about internship program. Zeynep-sensei is my project consultant, she helps me a lot with my project and she also wants to a publish research paper about my study. She will try her best to help me out in any circumstance even if I’m not a good researcher. I can tell that my impressive experience is the people I met here.

Chapter 6
Problems and Comments

Student

Problems
1. Language barrier.
2. Stubborn in problem.

Comments
1. Trying to learn Japanese.
2. Trying to ask for help.

Company

Problems
1. All project is large project.
2. Dormitory is expensive.

Comments
1. Need to adding some small project for short-term student
2. Trying to find a cheap one for student.

University

Problems
1. Too short period for attending internship program.
2. Too little scholarship.

Comments
1. Extend time for applying a program.
2. Give every student a scholarship.

References

Appendix A
Daily Reports

1st June 2018
Monden-sensei introduced me to Monden laboratory and show me my desk. He tells me that we will talking about my project next week so be prepare.

4th June 2018
Monden-sensei and Zeynep-sensei bring 3 projects to me and discuss about them. They let me choose one topic that I am interested in. I choose a topic about finding and improving graspable area of objects. Zeynep-sensei is a supervisor of this project, so she gives me some paper for understanding about the background of the project.

5th June 2018
After learning about the background of the project, she tells me about experiment and tell me to prepare myself for experiment tomorrow.

6th June 2018
I need to do first half of experiment for gathering gaze data in the afternoon with the help of Huang Nguyen, French internship student. I watch 6 videos with EEG headset on my head today. It really hurts.

7th June 2018
I need to do experiment again, today is final half of it. So, I watch 6 remaining videos then I finish gathering my data.

8th June 2018
After we finished experiment, Zeynep-sensei tells me that we will working on code next week, so this week she wants me to write a report about background of the project for check my understanding about what we are going to do next.

11th June 2018

I hand over my report to Zeynep-sensei, she says its seems like I understand about background of project now, so she gives me a standard code to work with and tells me to find IoU of the polygons in each object.

12th June 2018

After I find IoU, Zeynep-sensei tells me that we want to know relation between IoU and object size, so she tells me to do a background subtraction to find size of object. I start using matlib background subtraction

13th June 2018

After I use matlib background subtraction, I got RGB picture that hard to compute object size. I change my method to opencv background subtraction, it gives good results and we can now compute object size.

14th June 2018

When we have object size, Zeynep-sensei tells me that we will find relations between grasping polygons and object size, so she tells me to find polygon's size, distance between polygon centroid and center of mass of the object, and orientation of polygons relative to the object.

15th June 2018

I finished finding relations between grasping polygons and object size, but the graphs of relations are hard to analyze and get a value to use so I need to split object size into ranges (bins). Then I find mean value of each grasping polygon size respect to mean value of object size in each bin and plot error bars using standard deviation of each value.

18th June 2018

After I have mean value of each grasping polygon size, Zeynep-sensei tells me that now we can randomly generate our own polygon using normal distribution with mean value and standard deviation.

19th June 2018
Working on generating random polygons.

20th June 2018
Working on generating random polygons. Changing range of bins from 200 to 500 because some of the bins don't have any values, when we use 200.

21st June 2018
Finish with generation of random polygons. Next, I need to check performance of our generation algorithm by using IoU compare to ground-truth polygons.

22nd June 2018
I need to re-check the code because I get very low IoU values for our randomly generated polygons in comparison to ground-truth polygons.

25th June 2018
Working on re-code.

26th June 2018
After I re-code, I still get low IoU so Zeynep-sensei tells me that maybe it is because we using circular coordinates to randomly generate the polygons. It gives quite different values compared to ground-truth polygons. So, we need to change to an elliptic coordinate frame that fits better to the object than circular model. But, now she tells me to count gaze inside and outside polygon then prepare my mid-term project.

27th June 2018
I split gaze point into 2 group, gaze point of grasper and non-grasper. In each group I split gaze point into 4 group as gaze in positive polygons, in negative polygons, inside object (not in polygons), and outside object (not in polygons). I can see that some gazes are outside object but near the object, its mean sometimes people look at the edge

of an object then Zeynep-sensei says that we need to contour object to cover that gaze points but now she tells me that I need to focus on my mid-term presentation.

28th June 2018

Working on mid-term presentation.

29th June 2018

Working on mid-term presentation. Zeynep-sensei inform me about mid-term presentation informations.

2nd July 2018

Before mid-term presentation, I need to present to Zeynep-sensei first with my French friends. So, my mid-term presentation goes well. After presentation, my laboratory and yamauchi's laboratory have pizza party together.

3rd July 2018

After mid-term presentation, I start working with Internship report. Zeynep-sensei are busy about French student's paper, so I need to work on my own in this week.

4th July 2018

Working on report and presentation to my Thai teacher.

5th July 2018

I am working on re-code some function and present my project to my thai teacher.

6th July 2018

Zeynep-sensei wants me to check about very small size and very big size object with their foreground. And modify her function to make that function works.

9th July 2018
After I check on object and their foreground, I found that if object is very small, some detail will be missing on its foreground. Otherwise, if object is very big, some overuse detail will be appearing on its foreground like shadow. And If object color is like background, that part of object will be missing on its foreground. I also finish modifying her function.

10th July 2018
Zeynep-sensei tells me that we will start working with relation of object, gaze, and grasping polygons properties after we finish finding their properties from our elliptic model.

11th July 2018
Working on elliptic model.

12th July 2018
Zeynep-sensei ask me to try to generate grasping polygon by using gaze information as a centroid and make grasping polygon perpendicular to principal axis of elliptic model.

13th July 2018
After I finish generated grasping polygon, I evaluate performance of generate grasping polygon.

17th July 2018
I see that elliptic model gives better performance than circular model. This confirm that elliptic model is a right method for solve this problem.

18th July 2018
Now, we have object, gaze, and grasping polygons properties, we need to find their relation and build a model for each relation.

19th July 2018

Zeynep-sensei tells me that she will help me building a model for each relation, so now she wants me to work on my final presentation on next Monday.

20th July 2018
Working on final presentation.

23rd July 2018
Today is my final presentation day. My final presentation went well. All thanks to Zeynep-sensei for all help.

24th July 2018
After a final presentation, I need to finish my Internship report then I can give it to Zeynep-sensei to evaluate my report.

25th July 2018
Continue working on Internship report and Zeynep-sensei ask me to do some annotate data of my French friend.

26th July 2018
Continue working on Internship report and send an annotated data to Zeynep-sensei.

27th July 2018
Continue working on Internship report.

30th July 2018
Finish working on Internship report and send to Zeynep-sensei for evaluation of my report.

31st July 2018
My last day at Okayama University and I finish all document that are needed to be use in my Internship program.

Appendix B
Workplace Photos

13 July 2018

Activities: Last day of my French friend, Hoang.