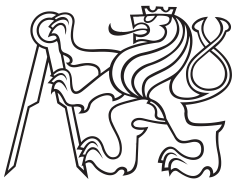
Master Thesis



Czech Technical University in Prague

F3

Faculty of Electrical Engineering

Part localization for robotic manipulation

César Sinchiguano

Supervisor: Dr Gaël Pierre Écorchard.
May 2019

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Dr Gaël Pierre Écorchard for his patient guidance and good predisposition to help me with my thesis.

Declaration

I hereby declare that the presented work was developed independently and that I have listed all sources of information used within it in accordance with methodical instructions for observing the ethical principles in the preparation of university theses. Prague, . May 2019

Abstract

The new generation of the collaborative robots allow the use of small robot arms working in an asynchronous or synchronous fashion with human workers. Such an example of the collaborative robot is the YuMi robot, dual 7-DOF robot arms designed for precise manipulation of small parts better known in computer vision as rigid body. For further acceptance of such robots in the industry, some methods and sensors systems have to be developed to allow them to pick parts without the position of the part being known in advance, just as humans do. This thesis is focused on the implementation of an algorithm for determing the posititon of the known parts. We first deal with a robot-camera calibration, then we propose a method to obtain the ground truth position of known parts. As step in between a 3D model of the known part needs to be created.

Keywords: manual, degree project, LaTeX

Supervisor: Dr Gaël Pierre Écorchard. Czech Institute of Informatics, Robotics, and Cybernetics, Office B-323,Jugoslávských partyzánů 3, 160 00 Prague 6

Abstrakt

Nová generace takzvaných spolupracujících robotů umožňuje použití malých robotických zbraní bez toho, aby byli izolováni od lidských pracovníků. Takovým příkladem spolupracujícího robota je robot YuMi, dvojitý 7-osý robot robotů určený pro přesnou manipulaci s malými částmi a dostupný v laboratoři Inteligentní a mobilní robotika CIIRC. Pro další přijetí takových robotů v průmyslu je třeba vyvinout některé metody a systémy snímačů, které by jim umožnily vybírat části bez předchozího znát umístění části, stejně jako lidé. Práce je zaměřena na implementaci algoritmu pro lokalizaci známých částí. Vedle lokalizace se část práce skládá z kalibrace kamery relativeley k robotovým a devolopingovým metodám pro získání pozemské pravdivé pozici dílů. ...

Klíčová slova: manuál, závěrečnná práce, LaTeX

Contents

1 Introduction **1**
1.1 Motivation 1
1.2 Goal 1
1.3 Thesis structure 2

2 Background **3**
2.1 something1 3
2.2 something2 4
2.3 something3 5

Bibliography **7**

Figures

2.1 An example of the social event on a webpage 5

Tables

Chapter 1

Introduction

Within this chapter, the reader receives an outline of the general context which surrounds this thesis. Starting with the motivation section and the ultimate goal to be accomplished, and a summary of the thesis' structure follow.

1.1 Motivation

For years. The industrial robot has undergoes through enormous development. Robot nowadays not only receives command from the computer. But also has the ability to make decision itself. Such abilities are well known in the world of the computer vision as recognizing and determining 6D pose of a rigid body (3D translation and 3D rotation). However, finding the object of interest or determining its pose in either 2D or 3D scenes is still a challenging task for computer vision. There are many researchers working on it with method that goes from state-of-the-art to deep learning means where the object is usually represented with a CAD model or object's 3D reconstruction and typical task is detection of this particular object in the scene captured with RBGD or depth camera. Detection consider determining the location of the object in the input image. This is typical in robotics and machine vision applications where the robot usually does task like pick and place objects. However, localization and pose estimation is much more challenging task due to the high dimensionality of the search in the workspace. In addition, the object of interest is usually sought in cluttered scenes under occlusion with requirement of real-time performance which make the the whole task even much more harder.

1.2 Goal

We attempt to provide an algorithm for determining the pose of a known parts similar to following pipeline "6D object pose estimation using RGBD data" [7]. In addition, a robot-camera calibration needs to be done, and a main requirement a 3D object model needed.

1.3 Thesis structure

The thesis consists of 5 chapters, **??** and **??**. The current chapter briefly describes the motivation and the goal for the part localization which we refer from here on through the whole thesis as 6D pose estimation of a rigid body in order to fit to the nomenclature giving in the perception field. Chapter 2 gives a background to camera calibration, openCV, open3D, ROS, prepossessing algorithm for segmenting the 3D image and related work about 6D pose estimation of the rigid body on which this work is building on. Chapter 3 describes the algorithms and the implementation for creating and collective ground data. Chapter 4 metric pair with the ground truth data. Chapter 5 concludes the thesis and showcase possible future works.

Chapter 2

Background

Information extraction (IE) is a complicated problem, and its goal is the automatic extraction of structural information from unstructured or semi-structured documents. Web Extraction (WE) is a particular case of IE where the corpus of documents are webpages. In the thesis, we are solving the problem of WE for the specific types of the data, namely the local (social) event extraction from a webpage.

2.1 something1

Besides the obvious application of WE in the search engine, there are also so-called vertical web services where WE is widely used too. Such websites allow to explore and work with the information on a specific topic. Such services usually do not produce the content by themselves, and rather they automatically collect the data from various sources, processes this data in some way and present to the user in the appealing form. Such services also called *aggregators*. A broad list of aggregators topics includes news, retail other advertisements, reviews, flight tickets, videos and pictures, recipes, social networks and many other. The list is very extensive, virtually for any subject discussed and presented on the Internet there are aggregators exist which try to grasp all other sources into one.

Aggregators are usually very popular because they provide the user with the big database, rich functionality, flexibility and instant updates, what helps a user to save time. Due to popularity of aggregator the 'original' content maker is becoming popular too and therefore the content maker has resources to produce new interesting content. The other side of this practice is traffic reduction on the original content maker websites. It happens for example if the aggregator is unfair and doesn't show the source of information.

The relationship between aggregators and original source makers today reminds the problem of chicken and egg. That's why search engines are very accurate with automatic extraction at the moment of showing the result for the search query. If Google shows you the correct answer right when you type the question, the website with this particular answer where Google took the

information from will lose the user. Since this site loses the user, it also loses monetizing and resources to produce the 'correct answers' in future; hence Google will lose the sources of correct answers. That's the reason why WE among Internet industry players is a controversial topic.

In the research area, WE is typically used in the tasks related to natural language processing and text mining. The reason for this is that the web page content is the semi-structured text mixed with the HTML markdown. Many research project centered around the analysis of various news sources as news articles, tweets, social networks, etc. Some of the possible tasks are sentiment analysis of news, topic recognition, summarization.

Some projects aim to build the system which collects huge amount of data, extract entities from them, determine relationships between them and my answer on human-language question (IBM Watson [**?**], Calais by Reuters [**?**], Wolfram Language [**?**]).

The extraction of data from the web page is closely linked with the Document Object Model (DOM) processing and implies the understanding of the web page structure from the 'browser point of view'.

In this way, WE is becoming a quite interesting, actual and challenging task. It includes both the technical background for the data retrieving and manipulation as well as theoretical background for the reasonable model building. This thesis includes all necessary parts of this task.

2.2 something2

In the thesis, we aim to create the system which automatically extracts the *event components* of the social event from a webpage. The social events might include musical concerts, meetings, performances, festivals, cinema and other social activities. Such events may be published on the different entertainment websites, but usually on a website of the event organizer.

To classify the entity on a web page as an event let's state that it must have the following *event components*:

1. The title
2. The description
3. The date and time
4. Certain location where event is taking place

In this thesis, we won't solve the problem of identifying if the event announcement is placed on the web page since it is the task from Information Retrieval field. We assume that the page already has an event and we want to

find and extract the structured information about it (i.e. extract four items above). On the picture 2.1 you can see a typical example of the page where four main event components are circled in red.

We consider *the webpage* as the rendered page where primary JavaScript already executed (it may be executed in future if the user interacts with a page). Such page has several important components: DOM tree, corresponding CSS, HTML code, resources (e.g. images) and rendered picture which the browser displays us.

For a webpage, we will collect all web elements from a DOM tree, and four of them would be components we need to extract. We will collect more than two hundreds features for every web element and build the model which distinguishes an event components from all other elements. In our problem we will consider four one-vs-all classifiers where each of them will detect associated event component.

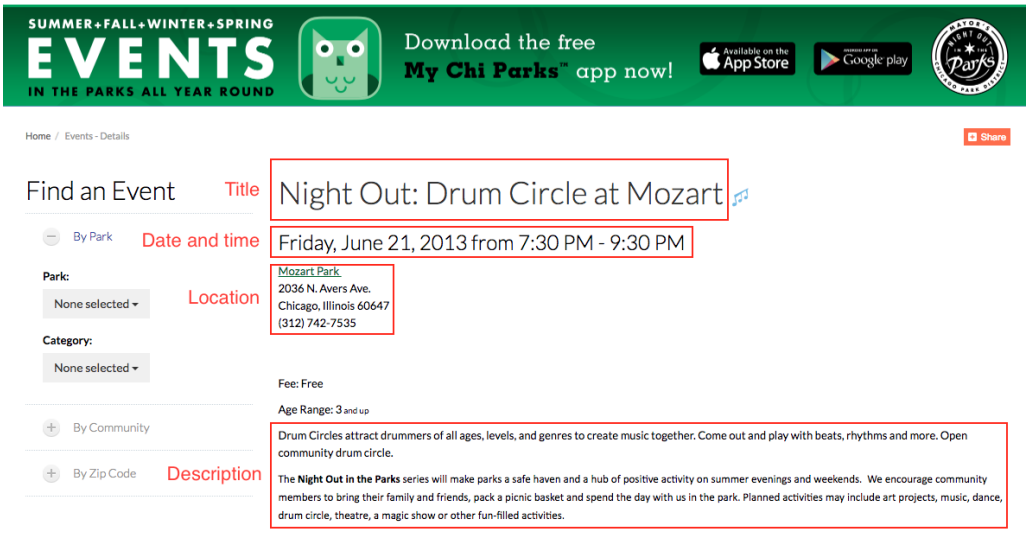


Figure 2.1: An example of the social event on a webpage

2.3 something3

The thesis consists of 8 chapters, **??** and **??**. The current chapter briefly describes the idea of the thesis and motivation for the task of web extraction. Chapter **??** covers necessary web concepts for this work: webpage, DOM, XPath, Microdata semantic markup, etc. Chapter **??** refers to an overview of state of the art Web Extraction techniques, overview of complementary problems from IE and related work. In Chapter **??** we explain the structure of our system together with the description of its main components. The method which we propose for the collecting of the training dataset will be

discussed in details in Chapter **??**.

Last three chapters are devoted to an iterative technical process of the dataset analysis. In Chapter **??** we will do several cleaning procedures in order to make the dataset reliable. In Chapter **??** we show the visualization and precise analysis of the extracted features and its relationships. In Chapter **??** we will build and evaluate several classification models. **??** will sum up our most valuable contributions and propose ideas for further work.

Lorep ipsum [6]

Bibliography

[1] J. Doe. *Book on foobar.* Publisher X, 2300.

[2] J. Doe. *Book on foobar.* Publisher X, 2300.

[3] J. Doe. *Book on foobar.* Publisher X, 2300.

[4] J. Doe. *Book on foobar.* Publisher X, 2300.

[5] J. Doe. *Book on foobar.* Publisher X, 2300.

[6] J. Doe. *Book on foobar.* Publisher X, 2300.

[7] J. Doe. *Book on foobar.* Publisher X, 2300.

[8] J. Doe. *Book on foobar.* Publisher X, 2300.