Memory Recall in Lifelog System Focusing on Memory Cues for Retrieving Events

LIANG JIE

Graduate School of Information, Production and Systems

Waseda University

A thesis submitted for the degree of

Master (Information)

July 2017

Thesis Supervisor: TANAKA Jiro

Abstract

Lifelog means using an intelligent wearable device to record human's life automatically. It is a way to change the human memory into digital data and it will be helpful to support memory recall when we take advantage of this recorded data. Our goal of this research is to create a lifelog viewer system which can help users recall memory effectively. To achieve this goal, we get help from several memory cues. We conduct 2 experiments and try to find out the important cues appeared in events remembered. After that, based on result, we design a system which can recognize these memory cues in lifelog data automatically and can support retrieving events with cue combination.

Keywords: Lifelog, Memory cue, Computer vision, Memory recall, Face recognition, Cues combination

Acknowledgements

First and foremost, I would like to show my deepest gratitude to Professor TANAKA Jiro, my supervisor for his guidance and advice in my research. His patience and encouragement help me overcome some difficulties.

I also want to thank all members in Interactive Programming Laboratory to give me a lot of help throughout my master student life. Besides, I would like to thank all participants who attend these experiments to provide data.

Finally thanks all my friends and family members for their support and accompany.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Lifelog means using an intelligent wearable camera which can take pictures usually every 30 seconds to capture human's daily life continually [1]. There are a lot of potential advantages of lifelog but so far to record past moment and help recall memory is still the core part.

1.1 Background

The vocabulary as "lifelogging" appeared from the 1980s [2]. Steve Mann did experiments with wearable computing to capture continuous physiological data. Since then, more and more work conduct on this filed. In the 2000s, lifelog camera with multiple sensors enter the market and become available to the public such as Vicon Revue [3]. In our research, we use the Autographer camera as our record device [4]. It was released in July 2013 and is used primarily for lifelogging, entertainment and travel.

1.2 Motivation

From ten years ago, memory recall was an important topic in psychology field. During our lifetime, we cannot remember everything happened so how to keep memory as much as possible is always most researchers interested in.

Recently there are numerous people research for memory recall in computer science field [5]. When lifelog becomes popular, to use this kind of device to turn memory into digital data in order to keep more memory is expected by lots of people. Of course there has been some related work but almost they focus on what is effective to memory recall. They do not take most advantage of the lifelog technology. We want to find out the significant benefits of using lifelog device and fully utilized this recorded data to keep everyone's past life.

1.3 The Organization of this Thesis

The rest of this thesis is structured as follows: In chapter2, we describe our research purpose and the approach we use to achieve this purpose. In chapter 3, we describe how we design our experiments conducted and discuss how to analyze the results we collected. In chapter 4, we discuss the problem of current viewer and how we improve it in our system. Based on the experiment results, in chapter 5, we describe the system design at first and describe a typical process when use our system. Next in chapter 6, we report how to implement our system in detail and the current limitation of this system.

Then in chapter 7, we introduce some related work to discuss the relationship with our work and compare what is the difference of our research from them. At last, in chapter 8, we conclude our finding results, our achievement and briefly discuss the limitation of our work and what can be improved in future work.

Chapter 2

Research Purpose

Memory recall is always a hot topic of study. With the advent of the lifelog, to store more memory has always been the target a lot of people strive for. In our research, the ultimate goal is to help users keep past memories.

2.1 Memory Cue

Usually, when we recall an event, we first think of several elements in this event. For example, who is involved, what I use or I see, when and where it happens. We call such elements which remind us recall memory cues. All events can be cued by this information [6, 7, 8, 9].

Apart from *who*, *what*, *when*, *where*, nowadays, more and more people spend considerable time on computer, smart phone to work or entertainment [10]. So we think *screen* also can be a potential memory and this idea first proposed in Matsuoka's paper [11].

2.1.1 Memory Cue Definition

In our research, for a better understanding of these five memory cues. We give them the following definition:

Who means the subjects related to this event alive in the world.

What means the specific objects we use or we see.

When means exact time.

Where means the specific place where the event happens.

Screen means whether users interactive with some display.

2.2 Research Goal

In general, there are two specific goals in our research:

(1) Find out which cues combination appeared in events remembered.

We want to know which cues combination can be possible to stimulate our memory. And from these cues combination, we want to find more effective combination. It means, with these cues, users can reminisce more past events.

(2) Create a system which can help users recall memory effectively.

After we know which cues combination are effective in memory recall, we want to create a new lifelog viewer system easy to use and reappear past scene for retrieving events with memory cues.

2.3 Approach

To achieve these goals, we first conduct 2 experiments. We design the experiment sheets and ask several participants to answer these sheets with the help of lifelog camera. And then by analyzing the results reflected from answer sheets, we can narrow down the cues combination to find out the effective ones.

Then based on the experiment result achieved, we design and implement a new viewer system for users to recall past memory. The system can recognize these cues information in lifelog data automatically and show originally scene in the form of pictures by retrieving cues.

Chapter 3

Experimental Setup

To find out cues combination used in memory recall, we design two experiments which consist of record part and recalling part. In record part, users need to use lifelog camera to take pictures. And then it comes to recalling part. We design the experiment sheet and there are several questions need to be answered.

3.1 Experiment design

3.1.1 Experiment goal

The goal to conduct our experiment is first to verify whether the lifelog can help users recall more events and quantify difference between using lifelog device and not using device and more important to find out which cues combination appeared in events remembered.

3.1.2 Experiment sheet design

In previous research, they found the following cues are effective in helping us recall past moments, it was: where, when, what, who [8, 11]. And in Matsuoka's paper, she thought screen may be important [11]. We want to confirm it and based on these 5 cues, we will introduce emotion level to describe the mental state of this event and event frequency to represent importance.

3.2 Experiment subjects

We have invited 12 participants(6 males and 6 females) to involve our experiments. They are all students in the Graduate School of Information, Production and Systems, Waseda University. Their ages range from 22 to 28 years(average age 23.92 years).

3.3 Experiment Device

We use a wearable lifelog camera named Autographer as our experiment record device.

Autographer is a hands-free, wearable digital camera. The users simply puts the camera on the clothes or wears on the neck and the camera will take photos automatically.

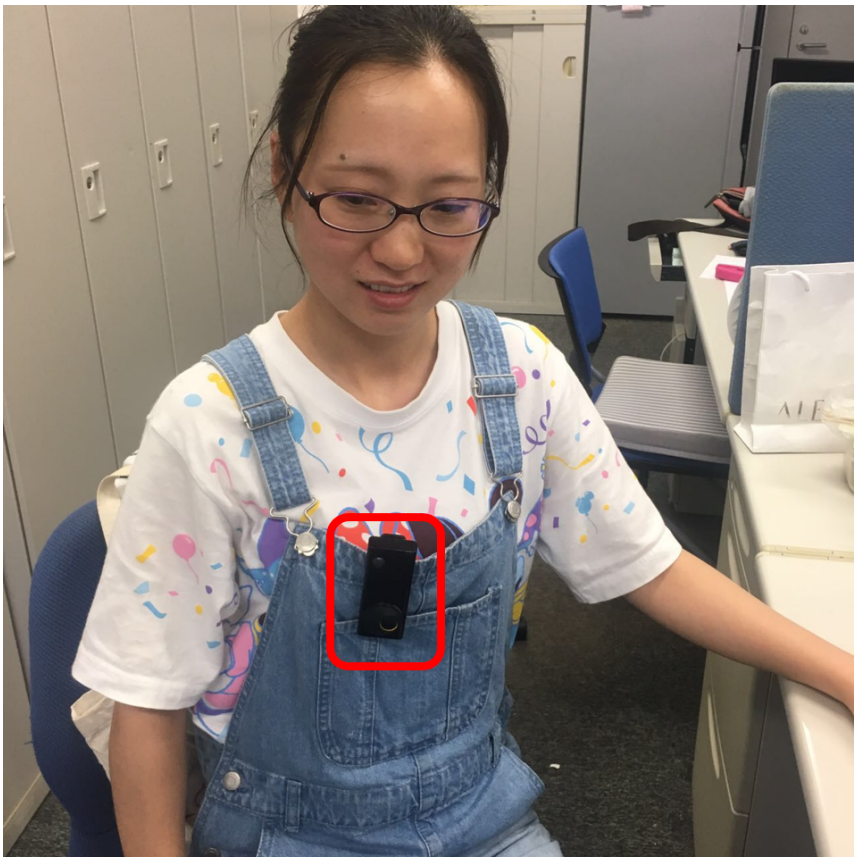

Figure 3.1: Autographer Camera.

Autographer camera's automatic shooting function is precisely calculated: its built-in 6 sensors, including GPS positioning, accelerator, light sensor, magnetometer, motion detector and thermometer to help determine when to shoot [12]. For example, when you run through the bus, the camera's accelerator will automatically sense and take pictures or when you go from a warm bar to the snowy streets to greet your friends, the camera also will automatically take a photo.

3.4 Experiment 1 procedure

Before experiment, we gather all these participants and did a brief explanation to them about experiment and how to use Autographer camera. And then the experiment conducted as below:

1. We divided all participants into 2 groups: each group consists of 3 females and 3 males.

2. Both 2 groups wore Autographer camera to collect their one-day life at ease for continuous 6 hours between 09:00 to 19:00. And they can switch off the device during their private time such as going to toilet.

3. Immediately after finish, for group 1, we gave them the sheet to write down the events they can remember by themselves not using any reference (Type A). And until they can think of nothing, we let them to view the lifelog photos with current viewer on computer. We first import their record data to the application and all pictures will show in time order. They can scroll to view the whole day pictures and then write down some additional events which can't remember by themselves. (Type B)

4. For group 2, they can directly view lifelog photos on current viewer and complete the sheet. (Type C). Each group should try their best to fill the sheet as much as possible.

5. After record, we collected these answer sheets from all participants.

3.4.1 Result

We have collected 196(63(Type A) + 34(Type B)+ 99(Type C)) sheets from 12 participants. By comparing Type A and Type B, it is obvious that

using lifelog camera can add us some more events. To compare A+B with C, we can find it's almost similar. It looks like independent thinking before using lifelog may not make sense.

Except for the number of recalled events, the appearance of important cue result is shown below.

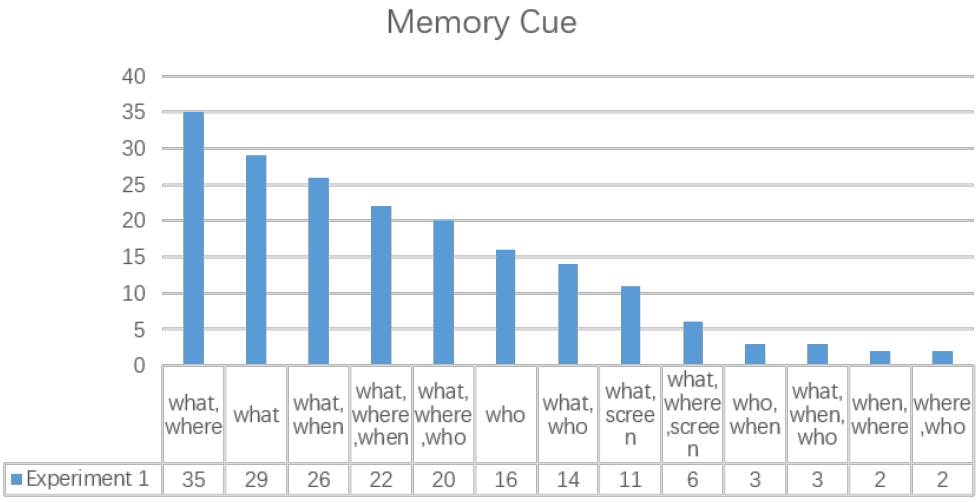

Figure 3.2: Important memory cue appeared in experiment 1.

From the figure, we can find that users usually write down multiple cues. More important is that what, where appeared most frequently and what is the strongest as a single cue.

3.5 Experiment 2 procedure

After a month, we gathered these participants and ask them to fill sheets again. The division of group and their type is same with the experiment 1. For example, For group 1 participants, they first recall what happened one month ago to fill the experiment sheet. When they can not think of anything, we show them the previous recorded data to help them reminisce more events and write down on sheets.

3.5.1 Result

We have collected 158(27(Type A) + 47(Type B)+ 84(Type C)) sheets after a month.

It is easy to understand after a month, the ability of memory recall dropped. The total number of recalled events decreased because some memory about what happened one month ago is missing.

We have put all collected data into a chart and compared with the previous result.

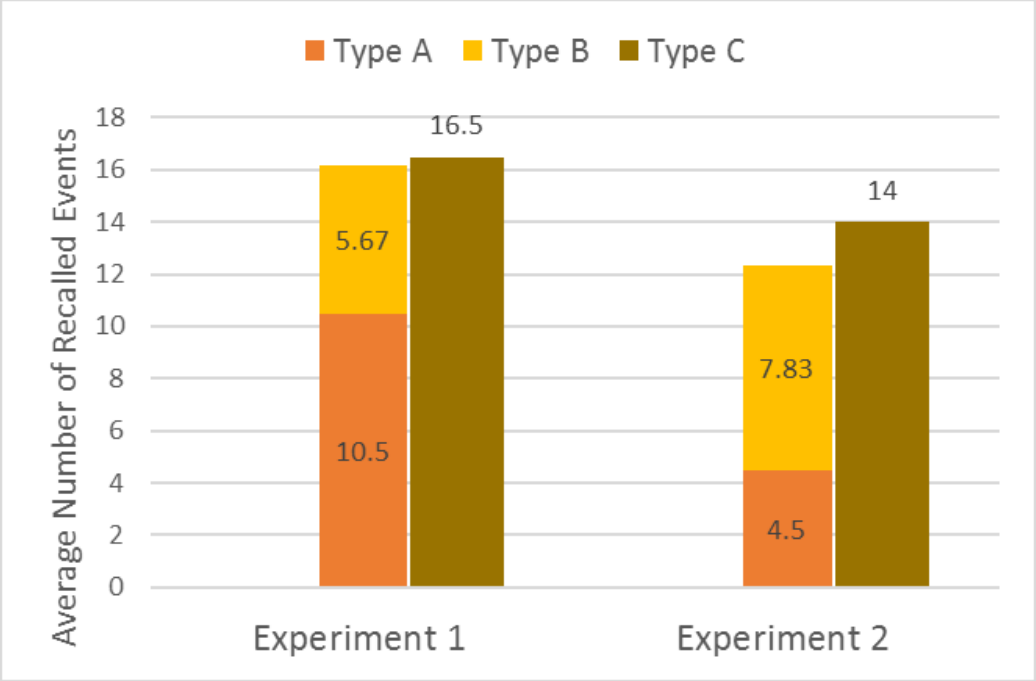

Figure 3.3: Comparison of the average number of recalled events by experiment 1 and experiment 2.

We can see from the Figure 3.3 that after a month the number of events recalled by independent thinking reduced but using lifelog can help us keep most of the memory.

By analyzing the important memory cues after a month, we found that there is no much difference with the experiment 1. Of course, to the same event, the answer of important cue may be changed but the combination of what and where still more important than others.

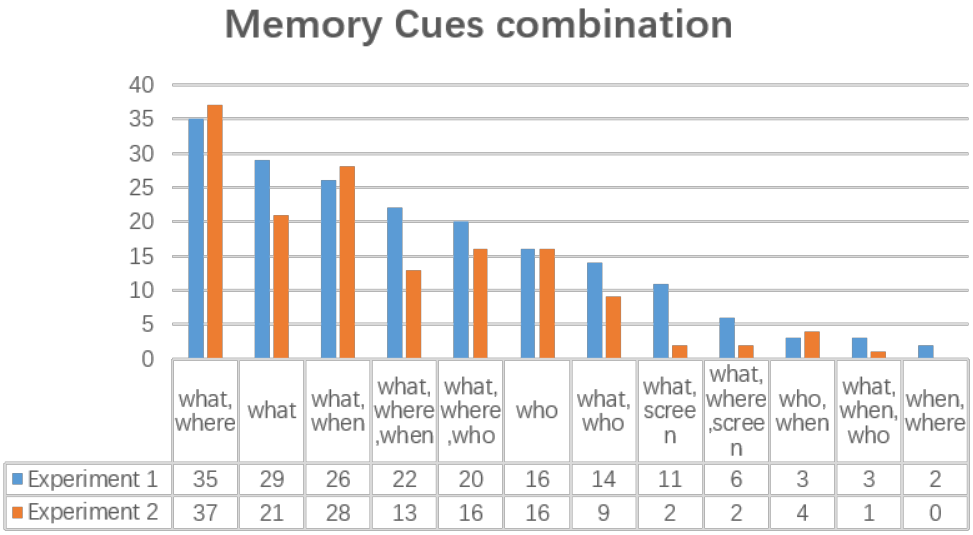

Figure 3.4: Memory cue result comparison in experiment 1 and experiment 2.

In summary, the results from 2 experiments shows that

1. By using lifelog picture as reference, it will support users more memory recall than independent thinking.

2. *What* is the strongest as a single cue, followed by *who*.

3. The combination of "*what* and *where*" is the most effective. And the second choice is "*what* and *when*", followed by "*what*, *where* and *when*", "*what*, *where* and *who*" and so on.

Chapter 4

Viewer Improvement Design

During our experiment, we found there are some limitations of current viewer. And we try to design a new viewer system to improve it.

4.1 Current Viewer

We observed how participants fill our experiment sheets using lifelog photos on current viewer and we found several problems of current viewer.

4.1.1 Typical memory recall process

1. Connect device to computer and import lifelog data to this desktop application.

2. All pictures will show in time order. Users need to scroll the mouse to view the whole day pictures.

3. Usually when users view some object, face or display they think of some event and fill the sheet and write down *what* or *who* or *screen* information.

4. Click a picture to view exact time to fill *when*.

5. Then they will watch the surrounding pictures carefully and try to find out the remaining information: *where* the location is.

4.1.2 Problem with Current Viewer

1) Picture is shown in time order.

2) *What* and *where* information is important but not directly supported.

3) Cue combination is not directly supported.

4.2 User Case

To solve these problems, we try to design a new viewer system. In this section, we will describe several typical user case to help users understand how to use our viewer better.

4.2.1 Import Images which recorded by user

After one day's record was finished, user can click to import all data into our system. The system will automatically process this meta data and the result will be show on the webpage such as faces, objects, location, creation time and so on .

4.2.2 View images via single cue

1. The user met somebody, and he cannot remember what happened with this guy. He can double click the person image to view the related images

with this person.

2. The user had an object in his memory, and he cannot remember the event scene. He can double click the object image to view the related images.

4.2.2.1 View images via multiple object cues

The user remembered several objects but he cannot remember the whole event. He can click all these object images and the system will retrieve the pictures which contain all these objects.

4.2.3 Query images via cues combination

1. The user used some object or saw some object somewhere. He can choose these objects and this location as cues combination and search the result.

2. The user used some objects at about 2pm but he did not remember the specific event. He can click to choose object as one cue and choose 14:00 as another cue, and the system will search the images between 13:30 to 14:30 which contain these object and show result.

3. The user saw some objects somewhere around 1pm, he wants to review these pictures. He can choose these objects, this location and click 13:00 to find all images during 12:30 and 13:30 satisfied the requests and recall what happened.

4. The user remembered someone with some objects somewhere, he can choose this face, objects and this place as cues combination. The system will search all images which contain the objects, person and fit the location information.

Chapter 5

System Design

Based on our experiment results, we design a new viewer system which can recognize object, face, location information automatically. Beside that, the system also support using cue combination to search directly.

5.1 System Front-End Design

In this part, we will introduce the design of the system front-end. The pages are listed as following:

a) Import Images Page

User can select the images which he wants to imported into our system.

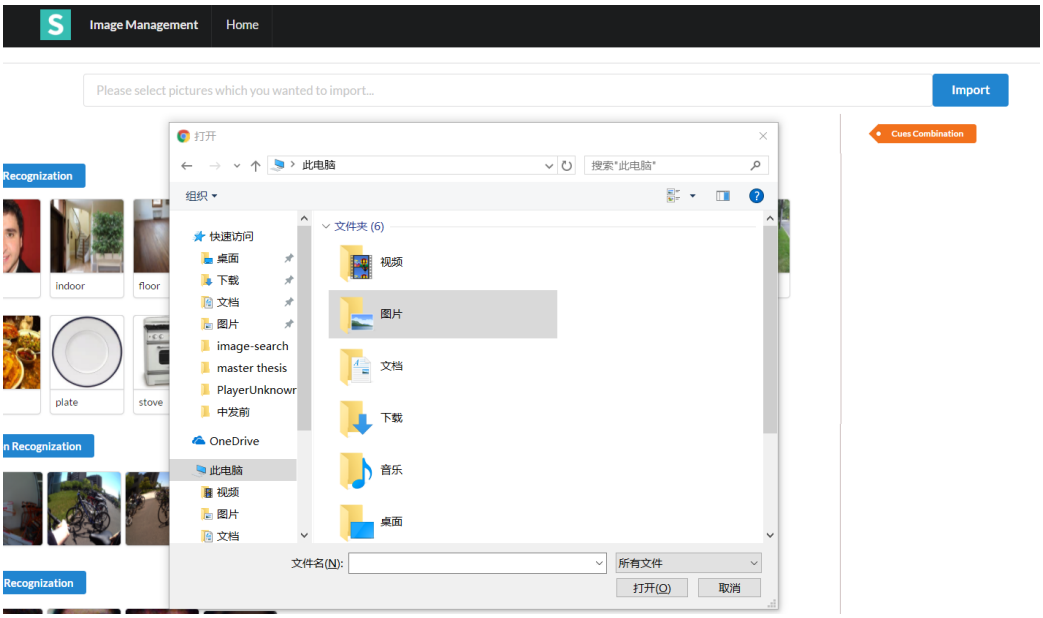

Figure 5.1: Import-Images-Page.

b) View Cues Images Page

After user imported these images into our system, we will handle the cues from these images and show the cues images in this page.

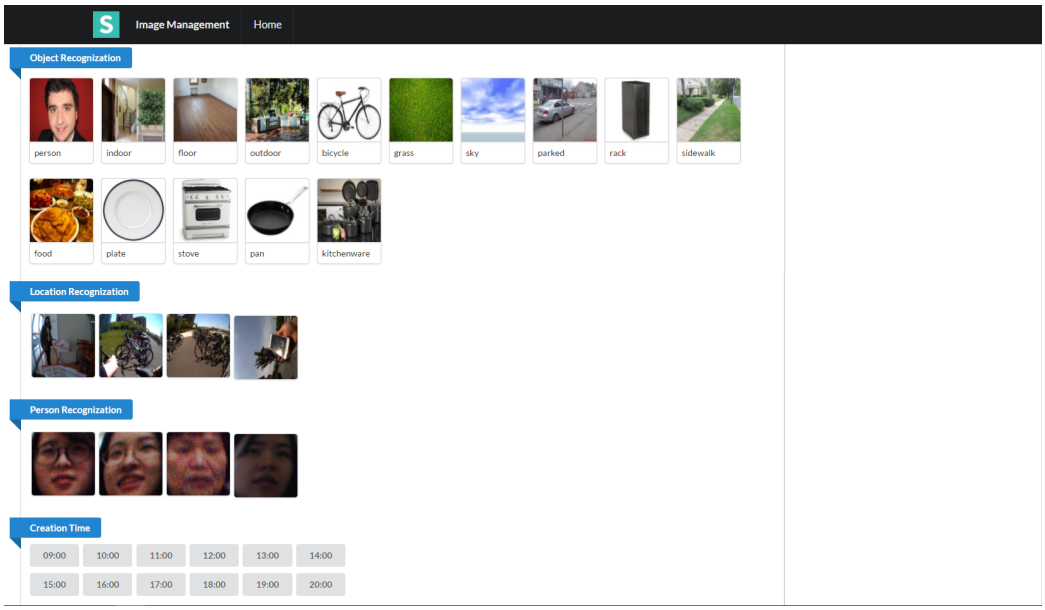

Figure 5.2: View-Cues-Images-Page.

c) View Images via Cue Page

When user double clicks the cue image, the system will show the images related this cue. And user can click next and prev button to view these images.

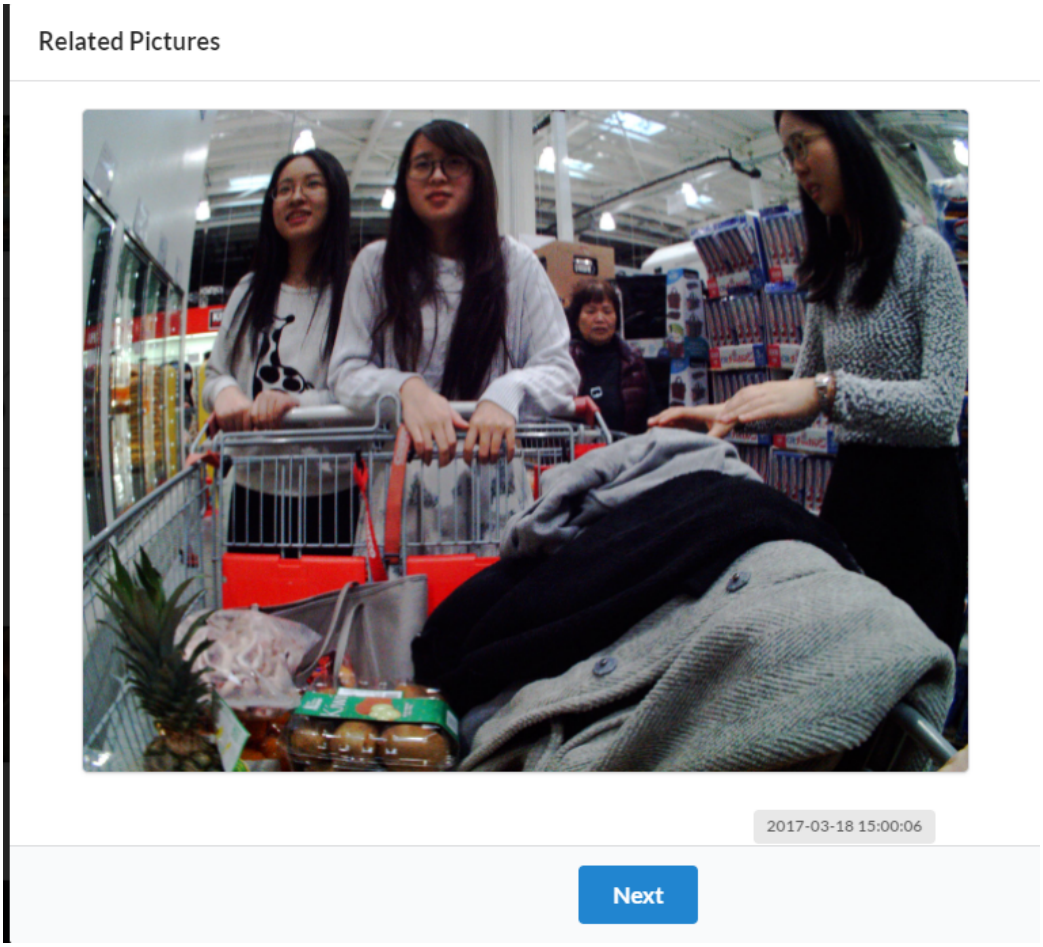

Figure 5.3: View-Images-via-Cue-Page.

d) Search Images via Cues Combination Page

If user wants to search images via cues combination, user can click the cue images and the image will show on the right of the page.

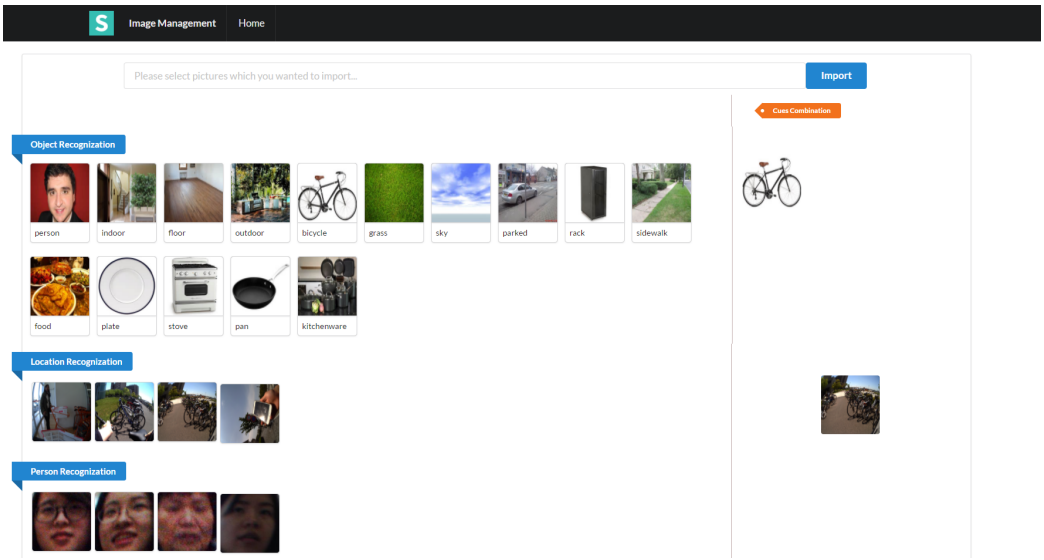

Figure 5.4: Search-Image-via-Cues-Combination.

e) View Result of Search Page

After user click the query button, we will show the result of the search via cues combination. User can click next and prev to view these images.

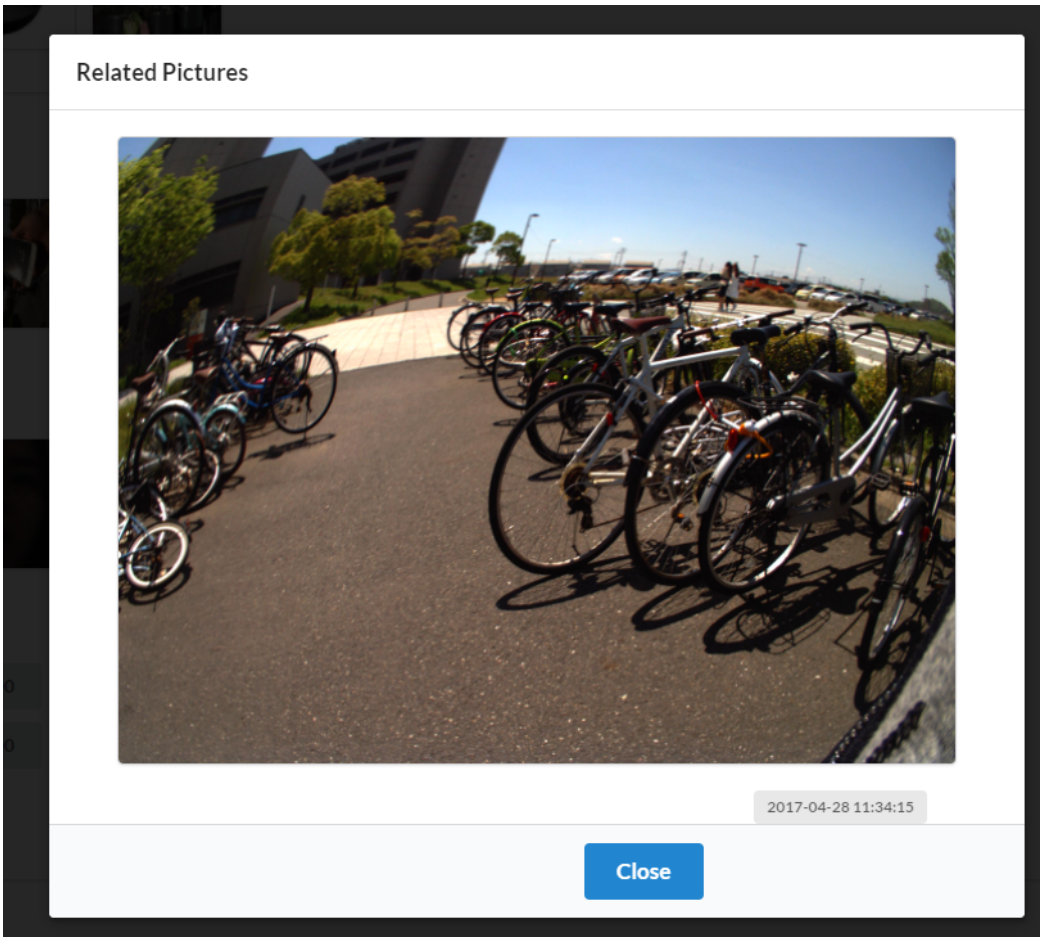

Figure 5.5: View-Search-of-Search-Page.

5.2 Typical memory recall process

We will describe a typical case to explain how to use our system in this part.

1. After one day's record was finished, users need to connect autographer device to computer and open our system website. They can click the

"choose-image" blank bar to choose the whole day's images for importing into our system and then click the "import" button. The system can automatically recognize this meta data, such as faces, objects, location, creation time information and you will see some preprocessing results on the webpage after a while. The users can view it easily.

2. When users recognize a human face, they can double click this face image here, it will show all original data contain this face in time order and the exact time information will be shown on each picture. You can click "prev" or "next" button to replace the image so users can write down all events related to this man by using these pictures as reference.

3. Users can also click to select picture as cue. For example, we remember a man and a bicycle. We can select his face, one bicycle image as cues. Of course we can also click the chosen cue to cancel it. After we choose several cues, we can click "Query" button to retrieve the result. In this case, the system will show the pictures both contain this person and this bicycle. When we see the original scene, we can easily write down the events.

4. If there is no result satisfied the requirement, the system will also send a feedback and you can choose another cue. So if there is something wrong with our memory, the system will help us to correct.

Chapter 6

System Implementation

In this chapter, we will describe how to implement the system in detail.

6.1 System Architecture

The system is designed using Browser/Server pattern. The advantage of this pattern is that we can use our system in any platforms which contains mobiles, pad and personal computer upon we deploy our system into the Internet. The system architecture can be designed as Figure 6.1.

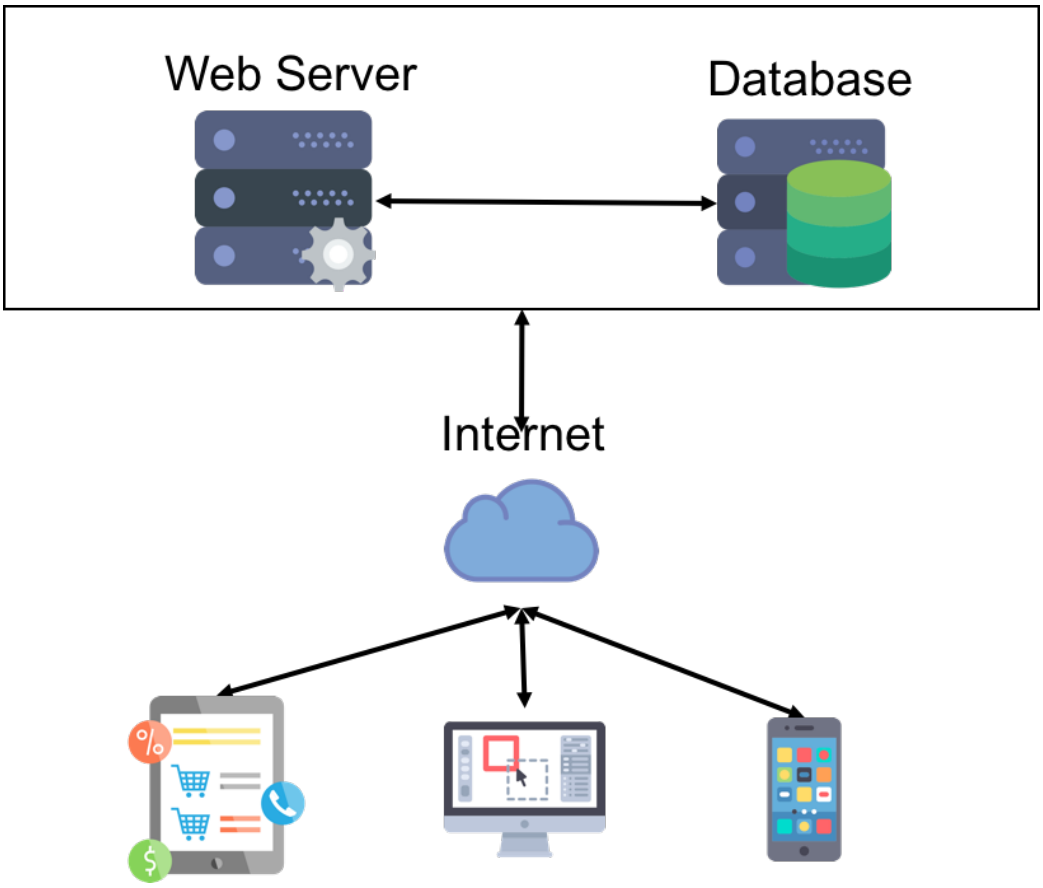

Figure 6.1: System Architecture.

From the picture, we can see that after we finished to develop our web server and database, and the deploy it on the Internet. We can reach our system in any platforms [13]. It's very convenience for user to use. The Browser/Server pattern also has other advantages are listed as below:

a) Operate anywhere without installing any specialized software

b) The expansion of the system is very strong

c) The maintenance and the logic of upgrade is very simple

For the first point, the reason is that we using our system via a browser, so if there is a browser in a device and our system is deployed on the Internet, the user can reach the website easily. As a result, the user could use our system without installing any specialized software. For the second point, we said the expansion is very strong because if we want a new feature in our system, we only need to update our web server. For the third point, the reason is similar with the second point. We have no client to maintenance. So we could spend less time on maintenance and upgrade. We just need to focus these on our web server.

We need to notice our web server in our system architecture. This part not only includes ourselves server, but also it includes the services from Microsoft and Google. We integrate the services from Microsoft and Google into our system. And the services include Computer Vision, Face Recognize, Bing Image Search and Google Maps Server. We use these services to handle the face, object and location in user's images.

6.2 System Framework

6.2.1 The design of the framework

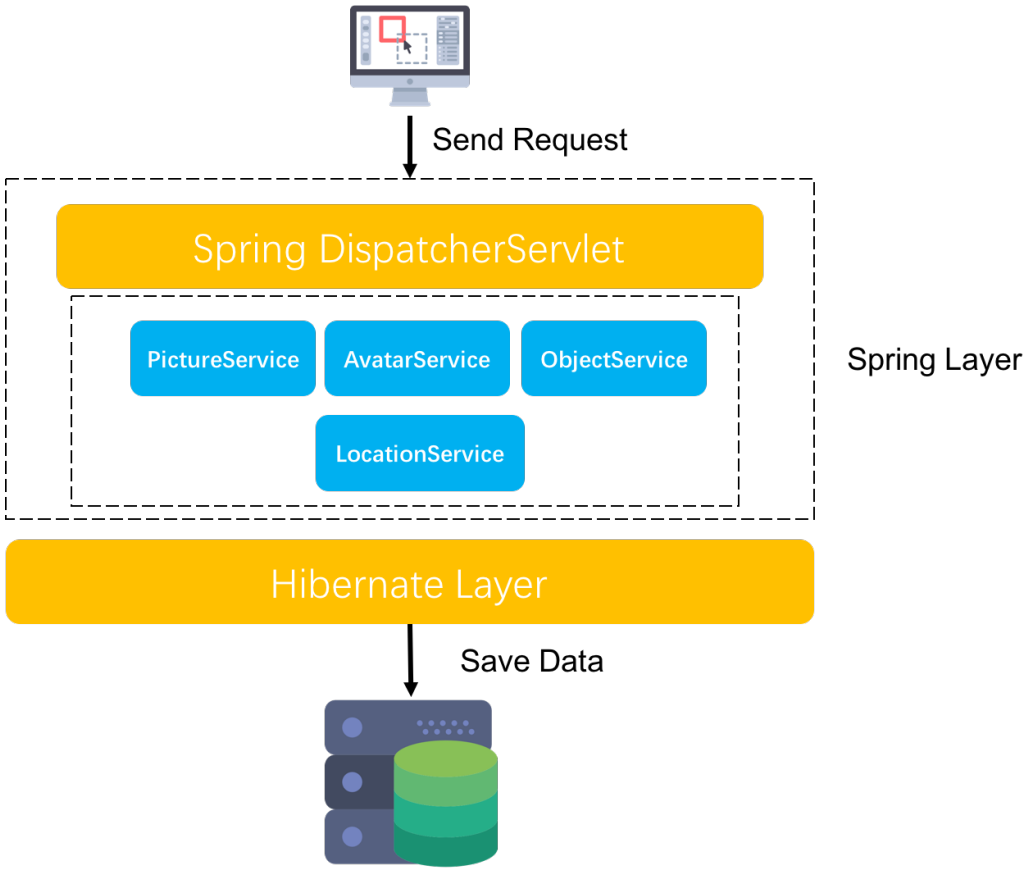

Figure 6.2: The Architecture of the Framework.

The implementation of the project mainly uses the combination of the Spring Boot and the Hibernate. The main purpose of using Spring Boot is that to realize the separation of the front-end and the back-end, and to reduce the coupling of the project [14]. The advantages of the Hibernate includes Ob-

ject/Relational mapping, JPA provider, idiomatic persistence, high performance, scalability, reliable and extensibility. The purpose of using Hibernate is to realize the separation of the business code and the data [15].

From the image we can see that when user send a request to our server via browser, the request will achieve to the Spring Dispatcher Servlet and the Spring Boot will dispatch this request to ourselves service. What's more, our service will get data and save data via Hibernate Framework. Then the server will return the data which we treated to the browser.

6.2.2 The design of the business code

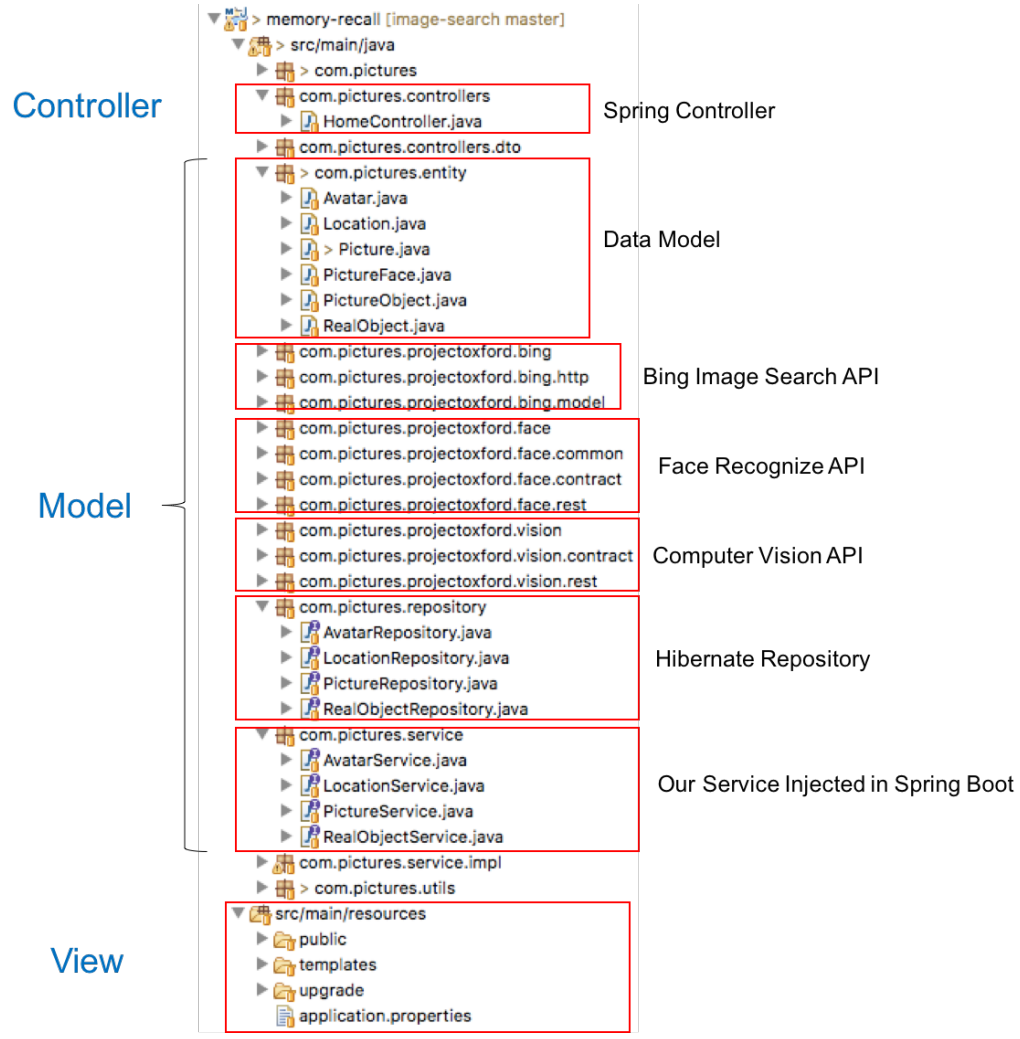

Figure 6.3: Code Architecture.

In our system, the design of the code follows the Spring MVC principles. MVC is a software design model, a business logic and data display interface, tissue isolation method code, business logic will be gathered in a component inside, and improvement in customization interface and user interaction at

the same time, do not need to write business logic. MVC is uniquely developed for mapping traditional input, processing, and output functions in a logical graphical user interface structure [16].

From the image, we can see that our system is made of model, view, controller. The model includes all the business logic of our system, such as Data Model, Services Injected, Bing Image Search, Face Recognize, Computer Vision and so on. The function of the Controller is to handle the requests from user's browser. Then we can treat the request using our service. The function of the View is that to show the page to the user. Using Spring MVC pattern can realize the separation of the back-end and front-end.

6.3 System Database

In this part, we will introduce the database design. It can be divided into two parts: the design of the tables and the relationships among the tables.

6.3.1 The design of the tables

We designed the tables for storing the information of the images which the user imported and finishing the work of viewing and searching. We designed six tables for our system. The details of the tables are listed as following:

a) Picture Table

This table will store the information of the images which user imported into our system.

Column Name	Type	Description
Id	LONG	The unique id of the image
Name	VARCHAR	The name of the image
CreationTime	DATE	The creation time of the image
CreationHour	INT	The hour of the image's time
CretionTimeString	VARCHAR	The creation time of the image
Longitude	LONG	The longitude of the image's location
Latitude	LONG	The latitude of the image's location
Location	VARCHAR	The location of the image

Table 6.1: Picture Table

b) Avatar Table

This table will store all the information of the faces image which recognized from the images which user imported into our system.

Column Name	Type	Description
Id	LONG	The unique id of the avatar image
Name	VARCHAR	The name of the avatar image
Uuid	VARCHAR	The uuid of the avatar image

Table 6.2: Avatar Table

c) Real Object Table

This table will store all the information of the objects image which recognized from the images which user imported into our system.

Column Name	Type	Description
Id	LONG	The unique id of the object image
Name	VARCHAR	The name of the object
Confidence	VARCHAR	The confidence of the object image
PictureName	VARCHAR	The picture name of the object
ParentTagsNumber	LONG	The number of the objects in the picture
BingImageUrl	VARCHAR	The url from the website of the object

Table 6.3: Real Object Table

d) Location Table

This table will store the location information of the image which user imported into our system.

Column Name	Type	Description
Id	LONG	The unique id of the location
Description	VARCHAR	The description of the location

Table 6.4: Location Table

e) Picture Faces Table

This table will store the information of the image's faces and the relationships between the image and the faces images.

Column Name	Type	Description
Id	LONG	The unique id of the face image
PictureId	LONG	The unique id of the image
Uuid	VARCHAR	The uuid of the avatar image

Table 6.5: Picture Faces Table

f) Picture Object Table

This table will store the information of the image's objects and the relationships between the image and the objects images.

Column Name	Type	Description
Id	LONG	The unique id of the object image
Name	VARCHAR	The name of the object
PictureId	LONG	The unique id of the image

Table 6.6: Picture Object Table

6.3.2 The relationships between the tables

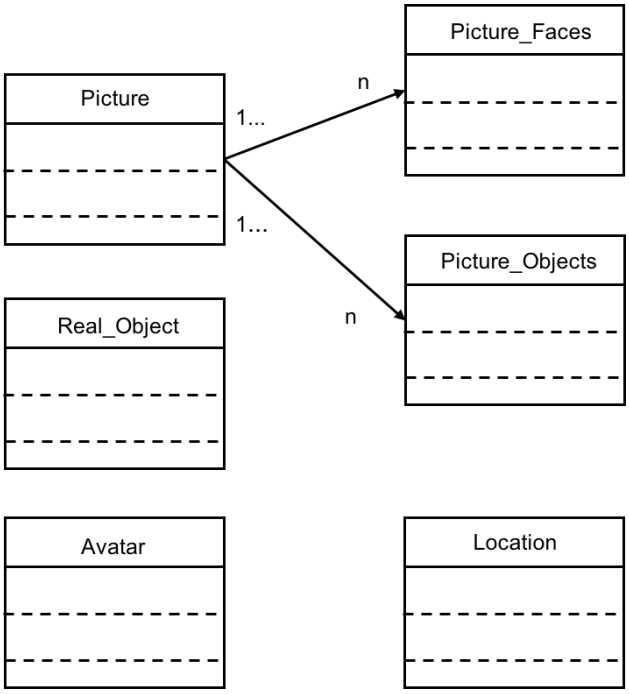

Figure 6.4: The Relationships of the tables.

From the image, we can see that there are 1 to n relationships between Picture and Picture_Faces, Picture and Picture_Objects. And the other is an individual table for the system. They are only store some useful information when we realize our system. The Real_Object Table will store all the objects in our system. The Avatar Table will store all the faces in our system. And the Location Table will store all the location information in our system.

6.4 Development Environment

6.4.1 The environment of hardware

The personal computer from labs: Lenovo Thinkpad X260

6.4.2 The environment of software

a) Platform

Windows 10 Operation System, Eclipse Neon, MySQL V5.7.18, Apache Tomcat V7.0

b) Running

Chrome Browser, Firefox Browser

6.5 API Introduction

6.5.1 Face Recognition API from Microsoft

a) Create a face list

Request URL:

http://westus.api.cognitive.microsoft.com/face/v1.0/facelists/{faceListId}

Request Parameters:

Field	Type	Description
faceListId	String	Id of the created face list
name	String	Name of the created face list

Table 6.7: Parameters of Creating Face List

The function of this API is that creating a face list which stores the faces we recognized from user's images. When we use this API, we need to provide a list id like 'Memory_Face_List_Id'and a list name like 'Memory_Face_List_Name'. And the API will return a response to show the result of creating the face list. If the response status is 200, that means we create the face list successfully [17].

b) Get a face list

Request URL:

http://westus.api.cognitive.microsoft.com/face/v1.0/facelists/{faceListId}

Request Parameters:

Field	Type	Description
faceListId	String	Id of the created face list

Table 6.8: Parameters of Getting Face List

Response Body:

Field	Type	Description
faceListId	String	Id of the created face list
name	String	Name of the created face list
persistedFaces	Array	Faces in the face list

Table 6.9: Response Body of Getting Face List

The function of this API is that we can get the face list we created using face list id. If we used the above API to create a face list which id is "Memory_Face_List_Id", we will use this face list id to get the face list. If

we send this request, we will get a face list which contains all the faces we recognize from the user's images [17].

c) Add a face into face list

Request URL:

http://westus.api.cognitive.microsoft.com/face/v1.0/facelists/{faceListId}/persistedFaces[?targetFace]

Request Parameters:

Field	Type	Description
faceListId	String	The id of the face list we created.
targetFace	String	The rectangle area of the face on the image.

Table 6.10: Parameters of Adding A Face

Request Body:

The bytes of the image

Response Body:

Field	Type	Description
persistedFaceId	String	The unique id of the face in the face list.

Table 6.11: Response Body of Adding A Face

The function of this API is that adding a face we detected from user's image into the face list which we created using above API. When we use this API to add a face into the face list, we need to provide the id of the face list which we created before and the area of the face location on the image. Then we will get a result of the action, the result contains the face id which

was stored in the face list which we created and we can get the face in the next time [17].

d) Detect faces from image

Request URL:

http://westus.api.cognitive.microsoft.com/face/v1.0/detect[?returnFaceId]

Request Parameters:

Field	Type	Description
returnFaceId	boolean	Return face ids of the detected faces or not.

Table 6.12: Parameters of Detecting Face

Response Body:

The bytes of the image.

Response Body:

Field	Type	Description
faceId	String	Unique face id of the detected face.
faceRectangle	Object	A rectangle of the face location on image.

Table 6.13: Response Body of Detecting Face

The function of this API is that detecting the faces on the image which user imported via our system. When we use this API, we need to convert user's image to bytes and set the 'returnFaceId'field to true. Then we will get the face unique id and the location of the face on the image [17].

e) Find the similar faces with the face

RequestURL:

http://westus.api.cognitive.microsoft.com/face/v1.0/findsimilars

Request Parameters:

Field	Type	Description
faceId	String	Unique face id of the detected face.
faceListId	String	The id of the face list we created before.
maxNumber	Number	The number of the top similar faces returned.

Table 6.14: Parameters of Finding Similar Faces

Response Body:

Field	Type	Description
persistedFaceId	String	The persisted id of the face in the face list.
confidence	Number	Similarity confidence of the candidate faces.

Table 6.15: Response Body of Finding Similar Faces

The function of this API is that finding the similar faces from the face list which we created to store our user's images. When we use this API, we need to provide a face id which we detect from the image, the id of the face list which we created before and the maximum number the similar faces returned. Then we will get the persisted id and the similar confidence of the faces [17].

Above all, we can use these API to finish the work of recognizing face from the image and query the images via a person's avatar.

6.5.2 Computer Vision API from Microsoft

In our system, we just use a part of the Computer Vision API. It is the API of analyzing image which can recognize the objects in the image. So we will only introduce the API which we use in our system.

a) Analyze Image

Request URL:

http://westus.api.cognitive.microsoft.com/vision/v1.0/analyze[?language]

Request Parameters:

Field	Type	Description
language	String	The name of the object which language return.

Table 6.16: Parameters of Analyzing Image

Request Body:

The bytes of the image

Request Body:

Field	Type	Description
tags	Array	The name and confidence of the objects in the image.

Table 6.17: Response Body of Analyzing Image

The function of this API is that recognizing the objects from the image which user imported into our system. When we use this API, we need to provide the language which we want to get and the bytes of the image. Then we will get the objects' name and the confidence from the server [18].

6.5.3 Bing Image Search API from Microsoft

The aim of using this API is just for enhancing to show the object image. Because when we record our daily life using the device, we will find there is too much objects in the image. If we just show the origin image on the website, user cannot ensure the image stands for which object. So we use this API to show the image which just contains the object.

RequestURL:

https://api.cognitive.microsoft.com/bing/v5.0/images/search[?q]

Request Parameters:

Field	Type	Description
q	String	The object name which you want to search.

Table 6.18: Parameters of Searching Images

Response Body:

Field	Type	Description
value	Array	Information of the images which contains the object.

Table 6.19: Response Body of Searching Image

The function of this API is that searching the images which only contains the object which you want. When we use this API, we need to provide an object name. Then we will get a lot of images which only contains the object and we can choose one to store into our database for the image of the object [19].

6.5.4 Google Maps API

This API is a library of maven module. So we just depend on it in our project. Our project is managed by maven, as a result, we just add the following code in our pom file:

```
<dependency>
    <groupId>com.google.maps</groupId>
    <artifactId>google-maps-services</artifactId>
    <version>0.1.17</version>
</dependency>
```

Figure 6.5: Google-maps-dependency.

When we import this library into our project, we will find there is a class file to get address via latitude and longitude. In our system, we follow the following steps to get the address from an image:

a) Convert an image into bytes

b) Load latitude and longitude from the image bytes

c) Use the function to get address: GeocodingApi.reverseGeocode()

d) Done. We already get the address from the image

6.5.5 Internal API

a) Import images into server

The function of this API is that helping user to import their images into our system.

Request URI:

/Images/upload

Request Method:

POST

Request Parameters:

MultipartFile[ ] images

The realization of this API:

1. Get the bytes of the image from the MultipartFile type
2. Save the bytes of the image into server's location
3. Handle the creation time of the image
4. Handle the location of the image
5. Recognize the faces in the image
6. Recognize the objects in the image
7. Save all the cues and the image into our database

b) Load avatar image from server to browser

The function of this API is that to show the avatar images for the users in our system. After user imported the images, we will show the avatar images which we deal with the images.

Request URL:

/avatars/download/id

Request Method:

GET

Request Parameters:

Long id

The realization of this API:

1. Get the avatar via the avatar id from database
2. Load the bytes of the image via avatar picture name
3. Return the bytes to the response body

c) Load object image from server to browser

The function of this API is that to show the object images for the users in our system. After user imported the images, we will show the object images which we recognized from the images.

Request URL:

/objects/download/id

Request Method:

GET

Request Parameters:

Long id

The realization of this API:

1. Get the object via the object id from database
2. Load the bytes of the image via object picture name
3. Return the bytes to the response body

d) Load location image from server to browser

The function of this API is that to show the location images for the users in our system. After user imported the images, we will show the location images which we get from the images via Google Maps API.

Request URL:

/locations/download/id

Request Method:

GET

Request Parameters:

Long id

The realization of this API:

1. Get the location via the location id from database
2. Load all the images which are in this location
3. Choose the first image in the image list
4. Load the bytes of the image using image name
5. Return the bytes to the response body

e) Load images via avatar id

The function of this API is to get the images which contain the avatar in our system. When user view the images in our system, user can double click the avatar image, then we will use this API to get the related images and show to user.

Request URL:

/avatars/id/images

Request Method:

GET

Request Parameters:

Long id

The realization of this API:

1. Get the avatar via the avatar id from database

2. Detect the face in this avatar via Face API

3. Find all the similar faces whose confidence is larger than 0.5 in the face list we created via Face API. This process will return the face UUIDs.

4. Load all the images via face UUIDs.

5. Convert the image to bytes, and return to the response body

f) Load images via object id

The function of this API is that to show the images which contain the object. When user view images in our system, user can double click the object image, and then we will call this API to show the images which contain the object. Request URL:

/objects/id/images

Request Method:

GET

Request Parameters:

Long id

The realization of this API:

1. Get the object via the object id from database

2. Load the image via object name

3. Return the bytes to the response body

g) Load images via location id

The function of this API is that to show the images which are in the location. When user view images in our system, user can double click the location image, and then we will call this API to show the images which are in the location.

Request URL:

/locations/id/images

Request Method:

GET

Request Parameters:

Long id

The realization of this API:

1. Get the location via the location id from database
2. Load the bytes of the image via location description
3. Return the bytes to the response body

h) Query images via cues combination

The function of this API is that to query the images in our system when user choose a cues combination. After user imported images in our system, if he wants to view images using cues combination, the system will call this API to show the images for user.

Request URL:

/images/query

Request Method:

GET

Request Parameters:

Field	Type	Description
faceIds	Array	The id collection of faces
objectIds	Array	The id collection of objects
locationIds	Array	The id collection of locations
times	Array	The time collection

Table 6.20: Request Parameters

The realization of this API:

1. Use the time to filter the images
2. Use the objects to filter the images
3. Use the lcoations to filter the images
4. Use the avatars to filter the images
5. Return the image list

6.6 The limitation of the system

There are two limitations in our system: the time of recognizing the image costs too much and the confidence value which we chose.

For the first limitation in our system, we cannot to improve it now. Because we use the Microsoft API to realize the recognition of the image. The

time is cost on the API call. So we cannot reduce the time cost now. But in future, the time cost will be reduced. For the second limitation in our system, we chose the confidence of the face similarity and the object confidence depends on our experiment. So the confidence value may be not a suitable value. But we can do more experiments in the future to get a suitable confidence value.

Chapter 7

Related Work

7.1 Memory Cues in Memory Recall

Since memory recall can be cued by "*who you met*", "*what you saw*", "*where you went*", recent research try to emphasize the importance of memory cues is not a few.

From 1978, Wagenaar spent 6 years on his study of his autobiographical memory [8]. By recording first and then recalling and scoring, he found "what he did", "when it happened" were both powerful as single cue and "the most successful orders elicited approximately 60% correct responses" were *what* came first, followed by *when*. The least successful order were *when* given first, followed by *who* or *where*.

7.2 Lifelog System

With the huge number of lifelog meta data, how to extract valid information and how to retrieve events more conveniently are nowadays researchers focusing on.

Matsuoka designed a system by linking lifelog photos with tags [11]. In her system, users need to manually select photos and edit tags of cues information one by one at first. Users can enter keywords to retrieve events and the system will show related pictures and suggest related tags as feedback. Users can also delete tags by themselves. This serious of operations increased the burden on the users and not friendly enough.

Memon developed a prototype system using an Android smart phone [20]. In his research, there are three key elements: nearby people, contextual objects and present location. His system can recognize people users have met instantly by using infrared sensors "which emits a unique infrared ID as an individual's identification" as a transmitter. With his system, lifelog related to current location or the specific location which users defined can be accessed directly. It is similar to object. He stored object images in database and the system may "read the textual information written on the objects to classify them." He also clarified using multiple key elements can retrieve more accurately.

Chapter 8

Conclusion and Future Work

8.1 Conclusion

Through the first experiment we conducted, we confirmed that using lifelog device can help users evoke more memory. By the way, independent thinking seems not so much important in the case of the presence of equipment. Of course, as our research goal set, the more significant conclusion is that we found users prefer to write down multiple cues when they recall past events and in which the most is the combination of "*what* and *where*".

With the supplement of experiment 2, we found there is no obvious difference compared with experiment 1 even though some memory is missing. Combining experiment 1 with experiment 2, the results revealed that *what* was the strongest one and *who* was also strong as a single cue. Moreover, in the case of more popular cues combination, "*what* and *where*" were the most effective, followed by "*what* and *when*", "*what*, *where* and *when*", "*what*,

where and *who*". However, what a little surprised is from our results, *screen* as a memory cue looks didn't has enough impact as we expected. We guess the possible reason is users spending time on computer or smart phone for entertainment and they think they waste time so almost people are unwilling to remember these things.

With our experiment discovery on memory cues appeared, we designed and then implemented a new viewer system in order to evoke users more memory by recognizing cues information in original lifelog data automatically. Our system sorted photos according by *what*, *where*, *who* information in stead of showing pictures in time order. For more accurate result, our system also support using different cues combination and multiple objects information to search directly. Users can view all related meta data retrieved one by one and they are organized by time order.

8.2 Future Work

We have created a prototype to realize these function but in practical application, our system will have the time problem because the process of dealing with the face information cost too much time. In future with the improvement of face recognition technology we hope our system can solve this problem.

The limitation of our system in recognizing part is we not providing *screen* cue so far. With current technology, it is still difficult for us detecting the

information appeared in screen but we believe in near future, it will has great progress and we can also complete our system at that time.

Appendix A

Experiment Explanation Document

Experiment Explanation

1. You should wear lifelog device to record your life for 6 hours. During that time, if you want to do some private things, you can turn off device temporarily such as going to toilet.
2. After finished, you need to return us the device and we will start the survey.
3. For group 1, you can only think by yourself. And we will give you some sheets. You should try your best to fill these as much as possible until you can think nothing. For the type on the sheet, you should write *A*.

   And then we will let you see lifelog pictures which you record on the computer viewer system, and you should fill other sheets again. This is Type *B*.
4. For group 2, you can directly fill the sheets with the reference of lifelog photos. And you should write type *C*.
5. To fill the sheets, you should write number in order from 1, and your name, this recording date in the form of Year/month/day (XXXX/XX/XX).

   Each sheet represents one event, and you should describe it in one word. This is called event description.

Who means the subjects related to this event alive in the world. If the person in this event is only you, you can write *myself*.

What means the specific objects you use or you see. It doesn't include the objects shown on screen.

When means time, in our case, you should write accurate to hour.

Where means the specific place where the event happens.

Screen means whether interactive with some display. You can circle one and if you choose ***other***, you should write down what it is. If there is no screen to use, you don't need to choose. Just leave it blank.

Importance is quantified by event frequency. There are 6 levels and the importance increases from 1 to 6. If you think it is very important, you can choose 6: 1/lifetime. If you think it is routine thing, not so much important, you can choose 1: 1/day.

Emotion means the following 6 feelings: pleasure, surprise, neutral, anxiety, sad, anger. You can choose 1 to describe your mental state of that event.

Emotional level is to describe the degree of this emotion. There are 4 levels and the strength is increasing from 1 to 4. If your emotion is neutral, you don't need to choose this emotional level.

Critical detail is some more specific detail instruction to this event. The question and the answer are all decided by yourself. You should ensure this part is closely related to the *event description*.

The last question is about the cues combination which you think important in this event. You can image and choose from 5 cues: who, what, when, where, screen. Of course you can choose one cue, 2 cues or even all 5 cues.

Appendix B

Experiment 1 sheet

Experiment 1

Number _____ Name _____ Date _____ Type _____

Event description (in one word) _____

Who _____

What _____

When _____

Where _____

Screen: TV computer smart phone other ( )

Importance

- ☐ 1 = 1/day
- ☐ 2 = 1/week
- ☐ 3 = 1/month
- ☐ 4 = 2/year
- ☐ 5 = 1/one year
- ☐ 6 = 1/lifetime

Emotion

- ☐ 1 = pleasure
- ☐ 2 = surprise
- ☐ 3 = anxiety
- ☐ 4 = sad
- ☐ 5 = anger

Emotional Level

- ☐ 1 = little
- ☐ 2 = moderate
- ☐ 3 = considerable
- ☐ 4 = extreme

Critical detail:

Question _____

Answer _____

Which cues combination you think is most important in this event?

_____

Appendix C

Experiment 2 sheet

Experiment 2

Number _____ Name _____ Date _____ Type _____

Event description (in one word) _____

Who _____

What _____

When _____

Where _____

Screen: TV computer smart phone other ( )

Importance

- ☐ 1 = 1/day
- ☐ 2 = 1/week
- ☐ 3 = 1/month
- ☐ 4 = 2/year
- ☐ 5 = 1/one year
- ☐ 6 = 1/lifetime

Emotion

- ☐ 1 = pleasure
- ☐ 2 = surprise
- ☐ 3 = anxiety
- ☐ 4 = sad
- ☐ 5 = anger

Emotional Level

- ☐ 1 = little
- ☐ 2 = moderate
- ☐ 3 = considerable
- ☐ 4 = extreme

Critical detail:

Question _____

Answer _____

Which cues combination you think is most important in this event?

_____

References

[1] Aiden R. Doherty, Alan F. Smeaton, Keansub Lee, and Daniel P. W. Ellis. *Multimodal segmentation of lifelog data.* In Large Scale Semantic Access To Content, p21-38, 2007.

[2] Steve Mann. *Wearable Computing: A First Step Toward Personal Imaging.* In IEEE Cybersquare, p25-32, February 1997.

[3] Lijuan Marissa Zhou and Cathal Gurrin. *A survey on life logging data capturing.* SenseCam Symposium, 2012.

[4] Katja C. Thoring, Roland M. Mueller, and Petra Badke-Schaub. *Ethnographic Design Research With Wearable Cameras.* CHI EA '15 Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, p2049-2054, April 2015.

[5] Shoko Kodama, Hideo Akaike, and Hiroyasu Kakuda. *Lifelog sharing system for memory recollection.* In IEEE 3rd Global Conference on Consumer Electronics (GCCE), p81-83, 2014.

[6] White R. T. *Memory for personal events.* Human Learning, 171-183, December 6, 1985.

[7] Mohsin Ali Memon. *Ensuring Privacy and Sharing Framework for Pervasive Logging.* Doctor Thesis, University of Tsukuba, 98 pages, 2014.

[8] Willem A. Wagenaar. *My Memory: A Study of Autobiographical Memory over Six Years.* In COGNITIVE PSYCHOLOGY 18, p225-252, 1986.

[9] Linton M. *Real world memory after six years: An in vivo study of very long term memory.* In M. M. Gruneberg, P. E. Morris, R. N. Sykes (Eds.), Practical aspects of memory, p69-76, 1978.

[10] John Rooksby, Parvin Asadzadeh, Mattias Rost, Alistair Morrison, and Matthew Chalmers. *Personal Tracking of Screen Time on Digital Devices.* CHI '16 Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, P284-296, 2016.

[11] Risa Matsuoka. *Understanding How Cues in Lifelog Affect Memory Recalls.* Bachelor Thesis, University of Tsukuba, 28 pages, 2016.

[12] Alexandra L. Young, Anya Skatova, Benjamin Bedwell, Tom Rodden, and Victoria Shipp. *The Role of Accidental Self-Reflection in Wearable Camera Research.* MobileHCI '15: Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct, P1062-1065, 2015.

[13] Linli Hu, Yuhao Wang, Dong Li, and Jing Li. *A hybrid client/server and browser/server mode-based universal mobile ticketing system.* The

2nd IEEE International Conference on Information Management and Engineering, p691-695, 2010.

[14] Tao Chen and Ronghua Ye. *Research on Data Persistence Framework Based on Spring Boot and MongoDB.* Computer Telecommunication, p71-74, 2016.

[15] Chuanlong Xia, Guangcan Yu, and Meng Tang. *Efficient Implement of ORM (Object/Relational Mapping) Use in J2EE Framework: Hibernate.* International Conference on Computational Intelligence and Software Engineering, p1-3, 2009.

[16] Praveen Gupta and M. C. Govil. *Spring Web MVC Framework for rapid open source J2EE application development: a case study.* International Journal of Engineering Science Technology, p1684-1689, 2010.

[17] Cognitive services face apis reference. `https://westus.dev.cognitive.microsoft.com/docs/services/563879b61984550e40cbbe8d/operations/563879b61984550f30395236/`. Accessed June 26, 2017.

[18] Cognitive services computer vision apis reference. `https://westus.dev.cognitive.microsoft.com/docs/services/56f91f2d778daf23d8ec6739/operations/56f91f2e778daf14a499e1fa/`. Accessed June 26, 2017.

[19] Bing image search apis reference. `https://dev.cognitive.microsoft.com/docs/services/56b43f0ccf5ff8098cef3808/operations/571fab09dbe2d933e891028f/`. Accessed June 26, 2017.

[20] M. MEMON, N. MAHOTO, F. KHUHAWAR, and J. TANAKA. *Retrieval of Life Logs based on Users Context.* Sindh University Research Journal. (Sci. Ser.) Vol. 47(4): 851-860, 2015.