

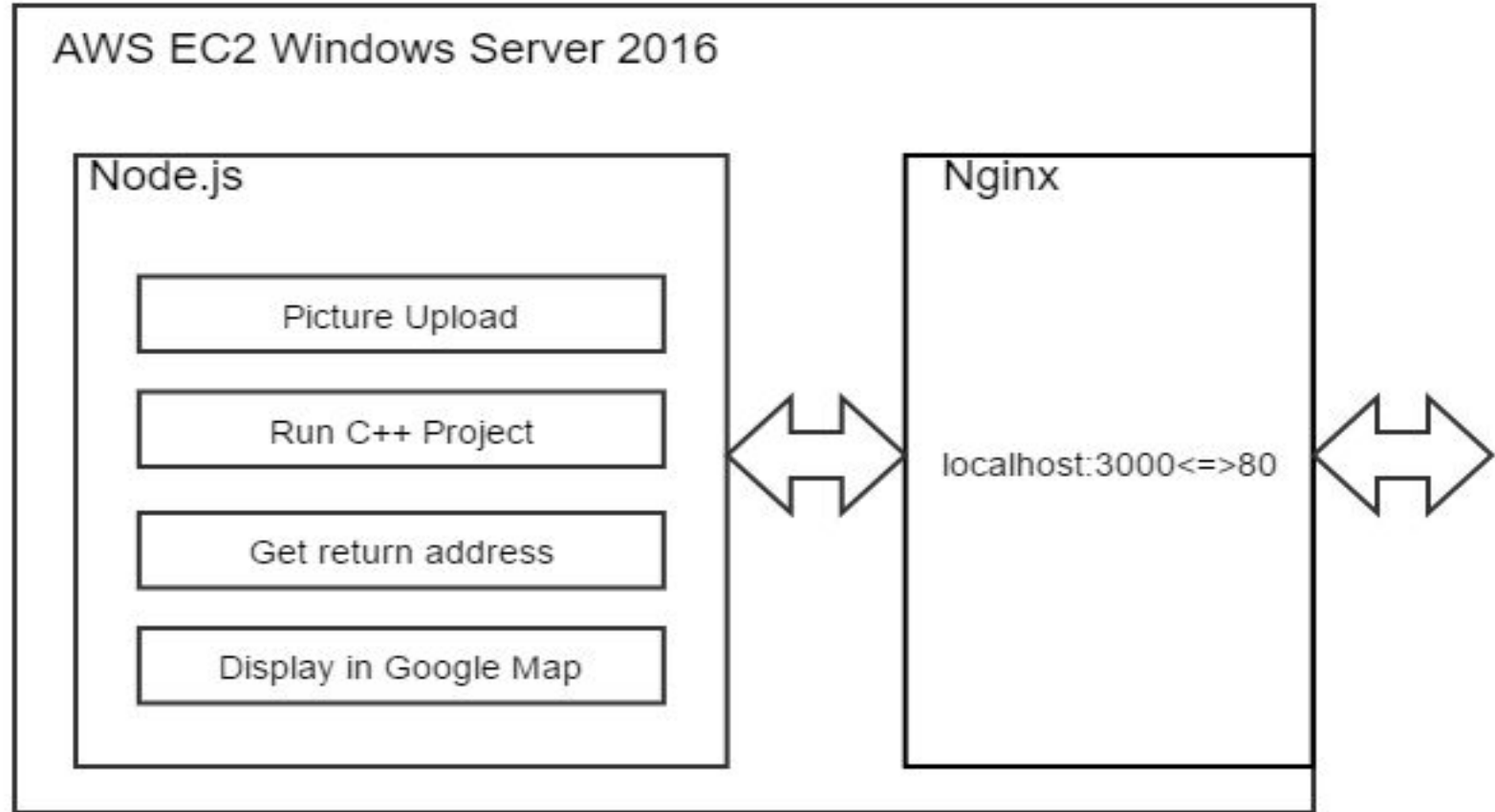
CLOUD BASED ADDRESS RECOGNITION SYSTEM

Using AWS, Node.js, VS2015

Xin DU 926753861

Zhixuan WU 219831020

PROJECT STRUCTURE



PICTURE UPLOAD

MULTER

A NODE.JS MIDDLEWARE FOR HANDLING
MULTIPART/FORM-DATA, WHICH IS PRIMARILY USED FOR
UPLOADING FILES.

MULTER

```
1  var express = require('express');
2  var path = require('path');
3  var multer = require('multer');
42 □ app.post('/profile', upload.single('avatar'), function (req, res, next) {
43    // req.file is the `avatar` file
44    // req.body will hold the text fields, if there were any
45  })
46
47 □ app.post('/photos/upload', upload.array('photos', 12), function (req, res, next) {
48    // req.files is array of `photos` files
49    // req.body will contain the text fields, if there were any
50  })
51
```

MULTER

```
15 var storage = multer.diskStorage({  
16   destination: function (req, file, cb) {  
17     cb(null, 'uploads/')  
18   },  
19   filename: function (req, file, cb) {  
20     cb(null, file.originalname)  
21   }  
22 })  
23  
24 var upload = multer({ storage: storage })
```

RUN C++ PROJECT APP.JS

```
22 var upload = multer({ storage: storage })
23
24 app.post('/upload', upload.single('image'), function(req, res) {
25   console.log('An image is uploaded...');
26   // Call 'process.exe' with the path of the uploaded image,
27   // get the output of process.exe, and
28   // return the result to the client.
29   console.log('process.exe ' + req.file.path);
30   console.log('process.exe ' + req.file.fileName);
31   var cmd = 'process.exe ' + req.file.path;
32   exec(cmd, function(error, stdout, stderr) {
33     res.send(stdout);
34   });
35 });
```

RUN C++ PROJECT PROCESS.CPP

```
1 #include <iostream>
2 using namespace std;
3
4 int main(int argc, char* argv[]) {
5     if (argc < 2) {
6         cout << "not enough arguments" << endl;
7         return 1;
8     }
9     // argv[1] is the path of target image
10    cout << argv[1] << " is processed" << endl;
11    return 0;
12 }
13
```

RUN C++ PROJECT INDEX.HTML

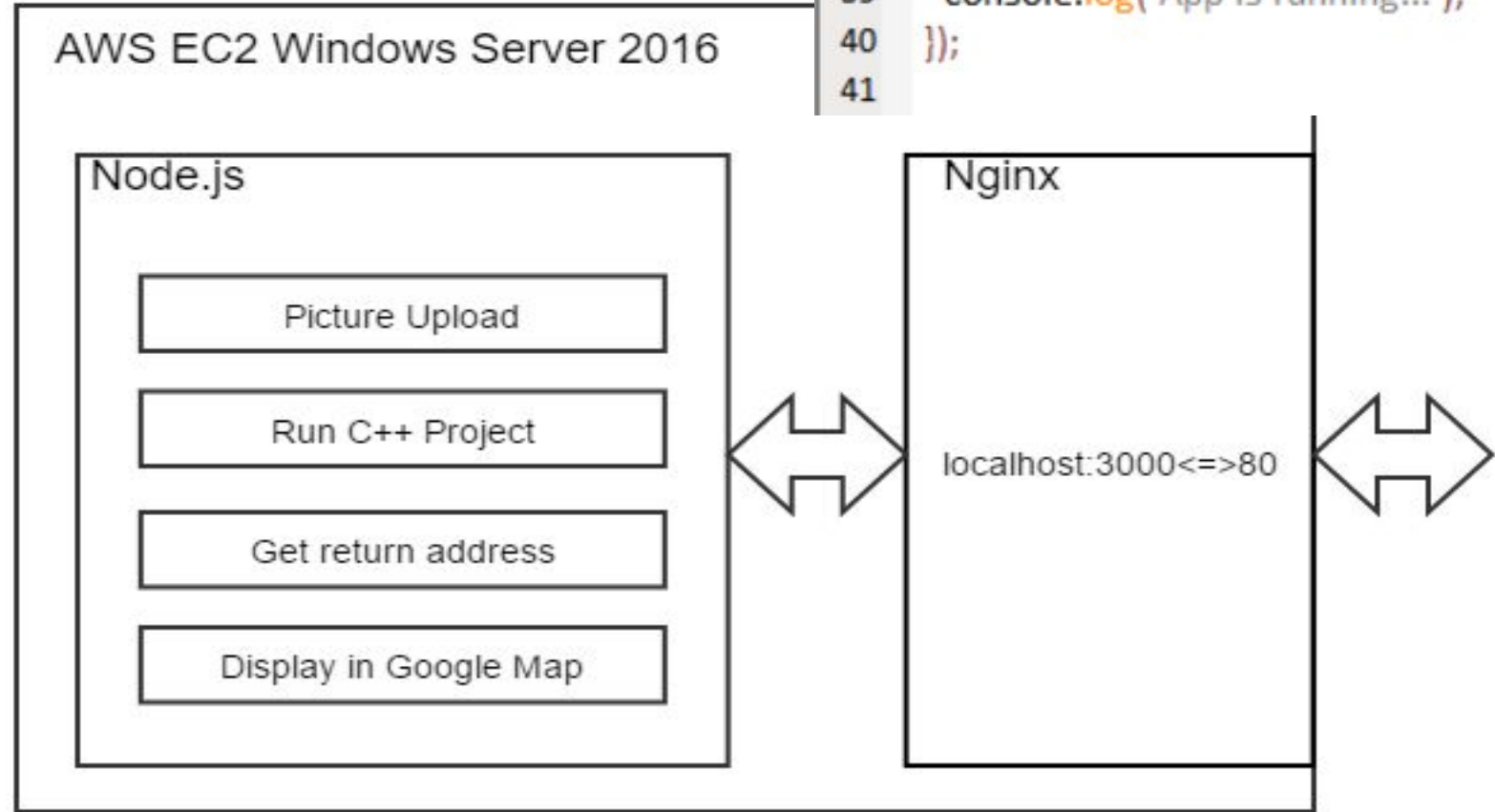
```
14 <body>
15 <input type="file" id="input"> <button id="upload">Upload</button>
16 <p id="output"></p>
17
18 <script>
19 $('#upload').on('click', function(e){
20     e.preventDefault();
21
22     var data = new FormData();
23     data.append('image', $('#input')[0].files[0]);
24 $.ajax({
25     url: '/upload',
26     type: 'POST',
27     data: data,
28     contentType: false,
29     processData: false,
30     success: function(response){
31         console.log('image uploaded');
32         $('#output').text('The image ' + response);
33     }
34 });
35 return false;
36 }
```

NGINX

A WEB SERVER, WHICH CAN BE DEPLOYED TO SERVE DYNAMIC HTTP CONTENT ON THE NETWORK.

NGINX WAS WRITTEN WITH AN EXPLICIT GOAL OF OUTPERFORMING THE APACHE WEB SERVER.

PROJECT STRUCTURE

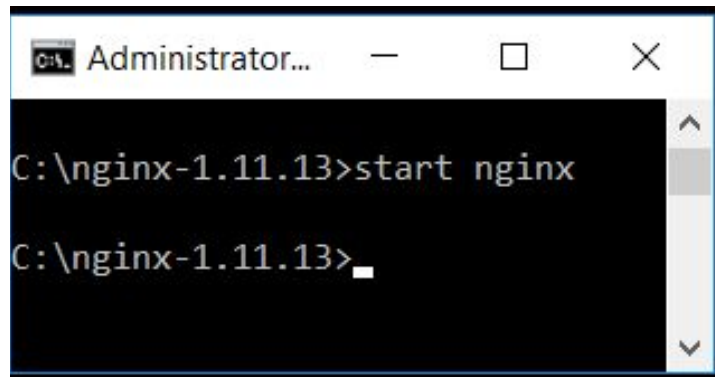


NGINX

nginx.conf - Notepad

File Edit Format View Help

```
server {  
    listen 80;  
    server_name ec2-54-172-0-231.compute-1.amazonaws.com;  
    location / {  
        proxy_set_header X-Real-IP $remote_addr;  
        proxy_set_header X-Forwarded-For $proxy_add_x_forwarded_for;  
        proxy_set_header Host $http_host;  
        proxy_set_header X-Nginx-Proxy true;  
        proxy_pass http://localhost:3000/;  
        proxy_redirect off;  
  
        # Socket.IO support  
        proxy_http_version 1.1;  
        proxy_set_header Upgrade $http_upgrade;  
        proxy_set_header Connection "upgrade";  
    }  
}
```



```
C:\nginx-1.11.13>start nginx  
  
C:\nginx-1.11.13>
```

TEST

<http://54.172.0.231>

DISPLAY IN GOOGLEMAP API

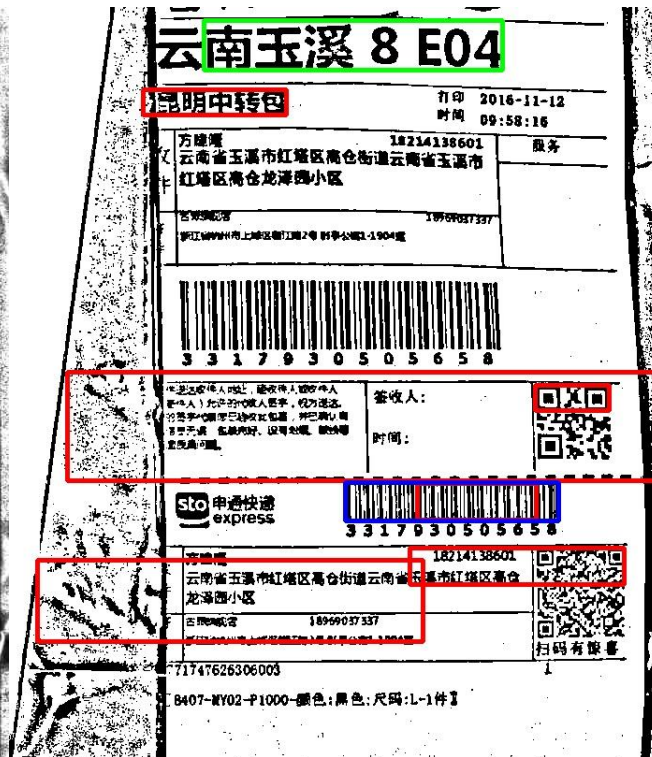
```
37 <body>
38   <div id="floating-panel">
39     <input id="address" type="text" value="上海市">
40     <input id="submit" type="button" value="Geocode">
41   </div>
42   <div id="map"></div>
43   <script>
44     header('Content-Type: text/html; charset=utf-8');
45     var markers = [];
46     function initMap() {
47       var map = new google.maps.Map(document.getElementById('map'), {
48         zoom: 11,
49         center: {lat: 43.04, lng: -76.14}
50       });
51       var geocoder = new google.maps.Geocoder();
52       document.getElementById('submit').addEventListener('click', function() {
53         geocodeAddress(geocoder, map);
54       });
55     }
--
```

DISPLAY IN GOOGLEMAP API

```
50
57 function geocodeAddress(geocoder, resultsMap) {
58     var address = document.getElementById('address').value;
59     geocoder.geocode({'address': address}, function(results, status) {
60         if (status === 'OK') {
61             resultsMap.setCenter(results[0].geometry.location);
62             var marker = new google.maps.Marker({
63                 map: resultsMap,
64                 position: results[0].geometry.location
65             });
66             markers.push(marker);
67         } else {
68             alert('Geocode was not successful for the following reason: ' + status);
69         }
70     });
71 }
72
```

ADDRESS RECOGNITION SYSTEM

WHAT'S THE PROBLEM AGAIN?



Province:

云南省

City:

玉溪市

TWO MODULE

OCR (Optical character recognition) MODULE:

Using OpenCV (Open Source Computer Vision) Library

Structurally describe a Picture/Frame (graphical)

Frame->List of Text Lines eg:



TextLine->List of Division eg:



Division->Matrix of Chars eg:



TWO MODULE

Address search module:

Using OCR module

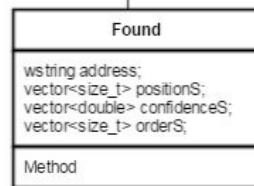
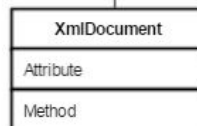
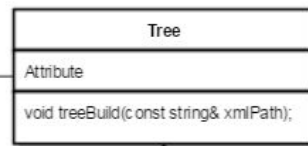
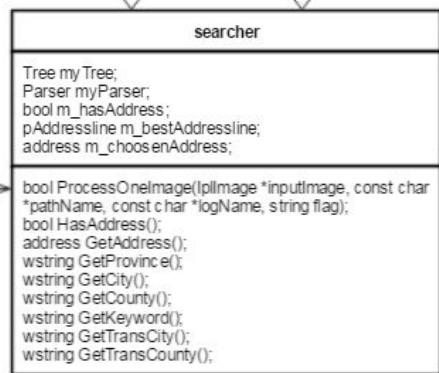
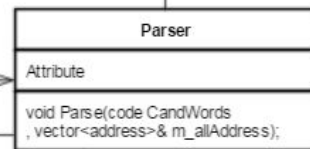
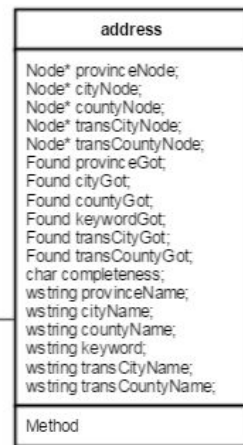
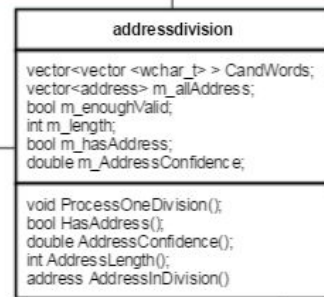
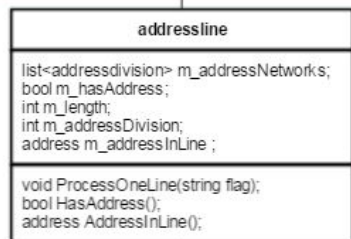
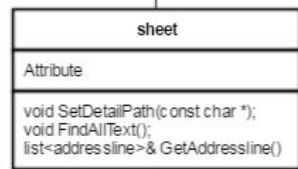
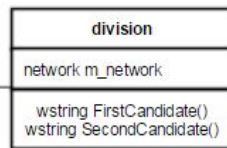
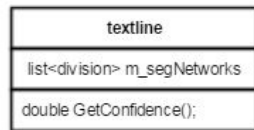
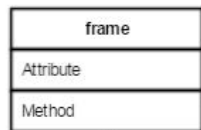
Structurally analysis info in Picture/Sheet (semantical)

Sheet->how many candidate lines have address?

line->address line or not? What's that?

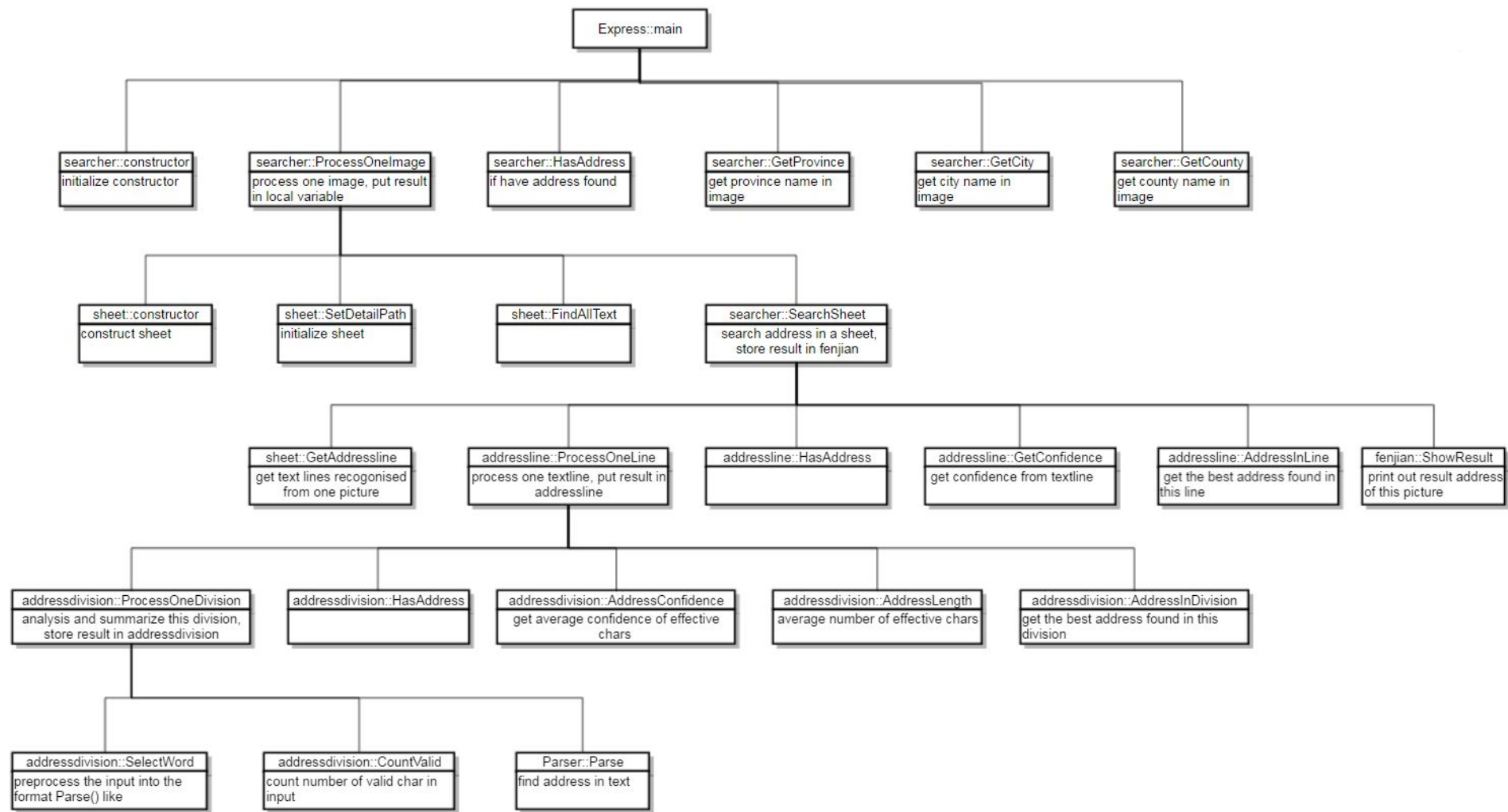
division->found anything make sense? How confident?

OCR_LIB



main()

SEARCH_LIB



PARSER STRATEGY

What to Find (only trust info can be cross checked)

At least to level of address on same branch

P+Ci+Co, P+Ci, P+Co, Ci+Co

Keyword + metropolitan

Transfer keyword + two address name in same province

Trimmed address 浙江萧山55 012 | 云南玉溪 8 E04

```
void SearchOnce(code& input, Node* key, size_t startPos, vector<Found>& nameGot);  
void SearchTrim(code& input, Node* key, size_t startPos, vector<Found>& nameGot);
```

```
void ExtractAddressWithProvince(code& input, vector<address>& allAddress);  
void ExtractCityPlusCounty(code& input, vector<address>& allAddress);  
void ExtractInTown(code& input, vector<address>& allAddress);  
void ExtractTransfer(code& input, vector<address>& allAddress);
```

TRANSFER EXAMPLE

sto 申通快递
express

标准
快递

成都转盐源24 005

集雅乐地区包

收 王阿凡 15282926640
四川省凉山彝族自治州盐源县盐井镇盐源县职业技术学校

寄 盐源 0571-89973505 15355072423
浙江省杭州市上城区甘王庙近江新村4-3-101

3 3 1 7 1 1 6 0 5 7 0 8 5

2016/11/0
01:00:52
9.210

2016/11/0
01:00:52
9.210

王阿凡 15282926640
四川省凉山彝族自治州盐源县盐井镇盐源县职业技术学校

寄 盐源 0571-89973505 15355072423
浙江省杭州市上城区甘王庙近江新村4-3-101

T: C211-#011-红色 L-1件

1

浙江德清德清县-野马路边宝

Prov:四川省 City:成都市

Trans keyword:转

Prov:四川省 City:凉山彝族自治州 County:盐源县

METROPOLITAN EXAMPLE

集 长春市市内包

收 孙明福 18686630968
吉林省长春市朝阳区前进街道吉林省长春市朝阳区前进大街湖畔成品9栋1602

寄 赵彬彬 0571-89973505 15355072423
浙江省杭州市上城区甘王庙近江新村4-3-101

3 3 1 7 3 4 1 1 6 9 6 0 4

2016/11/10 00:22:42
4/10

STC 申通快递 express

孙明福 18686630968
吉林省长春市朝阳区前进街道吉林省长春市朝阳区前进大街湖畔成品9栋1602

收 赵彬彬 0571-89973505 15355072423
寄 浙江省杭州市上城区甘王庙近江新村4-3-101

1: A409-MY-A409-MY21-刺绣款-深灰 L-件

迪克雅格旗舰店-shm hhd0309

Prov: 吉林省 City: 长春市

keyword: 市内

PARSER STRATEGY

How to judge(features caused by OCR)

Bigger chars preferred

Higher confidentiality preferred

More correct chars preferred

Increase of accuracy after new strategy added

60%→80%

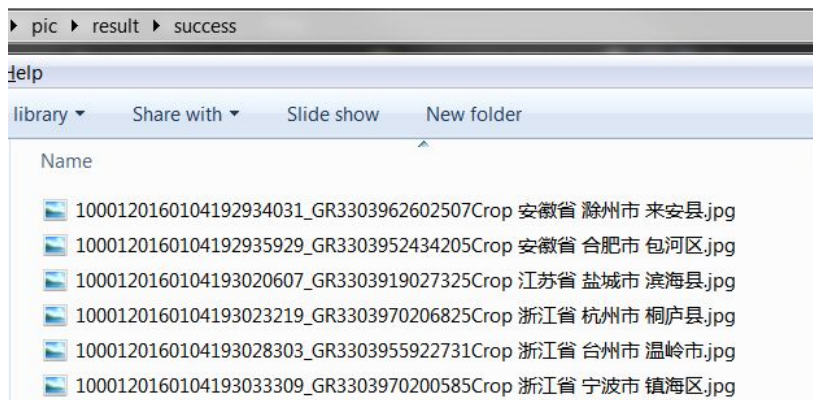
RELEASE PACKAGE

Codelevel123.txt: 1240 used chars
3Level123.xml: address tree

Name	Date modified	Type
pic	4/14/2017 15:41	File folder
3Level123.xml	3/19/2017 20:22	XML File
codeLevel123.txt	10/12/2016 12:38	TXT File
Express.exe	4/14/2017 15:29	Application
FenJianLib.dll	4/14/2017 15:29	Application extens...
LDA_total_RLF.dat	10/12/2016 12:38	DAT File
OcrBase.dll	4/14/2017 15:28	Application extens...
opencv_core240.dll	10/12/2016 12:38	Application extens...
opencv_highgui240.dll	10/12/2016 12:38	Application extens...
opencv_imgproc240.dll	10/12/2016 12:38	Application extens...
proto.pro	10/12/2016 12:38	PRO File
standard.dat	10/12/2016 12:38	DAT File

```
2  <root>
3  <北京市>
62 <天津市>
150 <河北省>
707 <山西省>
1101 <内蒙古自治区>
1102 蒙
1103 内蒙古
1104 <呼和浩特市>
1105 呼市
1106 呼和浩特
1107 <新城区>
1110 <回民区>
1113 <玉泉区>
1116 <赛罕区>
1119 <土默特左旗>
1123 <托克托县>
1126 <和林格尔县>
1129 <清水河县>
1132 <武川县>
1135 </呼和浩特市>
```

INTERFACE



```
Please input the image folder:
C:\Users\isaac\Desktop\Release\pic

Image 0, 1000120160104192934031_GR3303962602507Crop.jpg start!
found: 安徽省 滁州市 来安县

Image 1, 1000120160104192935929_GR3303952434205Crop.jpg start!
found: 安徽省 合肥市 包河区

Image 2, 1000120160104193020607_GR3303919027325Crop.jpg start!
found: 江苏省 盐城市 滨海县

Image 3, 1000120160104193023219_GR3303970206825Crop.jpg start!
found: 浙江省 杭州市 桐庐县

Image 4, 1000120160104193028303_GR3303955922731Crop.jpg start!
found: 浙江省 台州市 温岭市

Image 5, 1000120160104193033309_GR3303970200585Crop.jpg start!
found: 浙江省 宁波市 镇海区

Totally processed 6 images, and found 6 addresses!
```

**SPECIAL
THANKS**

TO: CHUNXU TANG