Abstract

Multiple image processing techniques has utilized to parse photos of a mall’s shopping instruction and a construct topological map of the mall. In this paper, we propose an indoor navigation system to address the problem. During navigation, we make use of this method to ﬁnd out the real time position of the user. Unlike most existing an indoor navigation system, which relies heavily on infrastructures and pre-labeled maps, our system uses only photos taken by cellphone cameras as input. With the help of a map and GPS, outdoor navigation from one spot to another can be done quickly and well. Moreover, we propose a new feature fusion method to help automatically identifying shops in a photo.

Introduction

These methods relies heavily on pre-installed infrastructures or special receivers, for example, RFID transmitters/ receivers , conﬁgured ﬂuorescent light receiver or many WiFi access points. Methods based on radio waves, magnetic ﬁelds, acoustic signals, or other sensory information collected by mobile devices has emerged. In this article, we propose a new vision-based indoor positioning system that replies on no other infrastructures but a camera. Ngrambased text classiﬁcation and scene recognition methods are used to recognize shop in real-time environment. DLA methods are used to extract information from the map provided by most shopping malls. The ﬁrst module is an image parser, where images of shopping instructions (ﬂoor plan and shop list) are parsed and topological map of the shopping mall is constructed.

Image Processing

In preprocessing step, texts/icons are detected, then recognized and ﬁnally removed from map image so that topological information can be extracted correctly later. Icons and texts provides semantic information about the map, while colored blocks provide topological information. A shopping mall ﬂoor plan typically contains a set of icons, texts and colored blocks.

Text Detection and Recognition

Therefore, texts can be extracted by detecting maximally stable regions. Following the method described in (Li and Lu 2012), we detect MSERs in the image as CCs. Detected CCs are ﬁltered based on their size, eccentricity and aspect ratio. In our practice, we locate texts using MSER-based method mentioned above. Then we crop text areas from the image and deliver them to tesseract software and get its recognition result. For text recognition, we use the open source tesseractocr software package. It performs OCR in a bottom-up manner, where CCs are ﬁrst detected independently and then clustered together to form text blobs and lines. We regard texts as corruptions and apply in painting method. In painting keeps color blocks in the image smooth so that later a segmentation operation can be done correctly to separate accessible area and different shops.

Text Removal

In painting keeps color blocks in the image smooth so that later a segmentation operation can be done correctly to separate accessible area and different shops. We regard texts as corruptions and apply in painting method. For simplicity purposes, we will refer to the image with texts and icon removed as image R.

Floor planning Parser

Based on this presumption, image binarization and CC analysis is used to identify accessible area. Accessible Area Extraction In a ﬂoor plan, accessible area (AA) usually has an intensity different from any component else so that viewers can easily ﬁnd out where they can go. In this step, accessible area and shop blocks are separated and labeled. Accessible area and shop blocks are separated ﬁrst. Finally the accessible area is segmented into different nodes.

Accessible Area Extraction

After this, we deﬁne the CC corresponding to accessible area (denoted as CCA) as: CCA = argmin(E(CCi)),CCi ∈ CC (1) where E(·) denotes the Euler number . Based on this presumption, image binarization and CC analysis is used to identify accessible area. In most cases, shop blocks are surrounded by accessible areas, resulting in many ’holes’ in it and its Euler number being the minimum. An Euler number is the number of CC object minus the number of holes in the object.

Shop Block Segmentation

The stopping condition for ﬂood-ﬁll is that the difference between current pixel and last pixel is greater than a threshold value. We choose 25 as the threshold in our experiment. This method is based on the presumption that different shops are separated by curves with distinguishable intensities.

Accessible Area Segmentation

In order to perform navigation, the accessible area needs to be further segmented based on the different attributes of its pixels. An indirect weighted graph is constructed to represent the topological layout of the ﬂoor plan. Figure 3 shows the result of AA segmentation and how the graph is constructed. The weight of an edge is the distance between centroids of two nodes on the ends of the edge. In our system, we treat each pixel of the AA as an observation spot.

Shop List Parsing

In real life, a user is interested in the name of a shop rather than its ID number on the ﬂoor plan. Therefore, it is important to map shop IDs with shop names. A shop list is usually actable-like document with each column being either shop names or shop ID numbers. In this way, shop names and IDs can be paired correctly. One of the two columns should be shop names and another column should be the shop IDs. There may be several table blocks in a shop list. To pair separated shop name and shop ID together, the result of XY-cut is stored as a tree, where each leaf nodes represents an XY-cut block. Each paired block is then splited into different lines with each line being a mixture of a shop name and a shop ID. In each recursion, we dichotomize the block in X direction and cut Y direction into several parts. Finally the salient map is cut into several XY-cut blocks, with each of them being a column of shop names, shop IDs or a mixture of both.

Map Construction

The name of a shop is mapped with its ID through analysis of shop list, and the IDofa shop is mapped with its blocks in map through analysis of ﬂoor plan.

Environment Sensing and Localization

Decorations or skewnesses made deliberately by designers can cause trouble to OCR software which are trained on standard printing fonts. Secondly, motion, luminance and visual angle variance can make logo texts hard or impossible to detect and/or correctly recognize. 2014) and, we show in experiment that using pure ngram for shop recognition can still fail in many cases. The most severe problem with encoding a shop name using ngram is the possible ambiguity when shop logo text is incorrectly or only partly captured. If such style can be captured and utilized jointly with ngram feature in the recognition process, we may expect some gain in the accuracy. And has been proved to be useful for shop recognition. This method is inspired by the fact that many famous brands have their shops decorated in a uniﬁed style. In this section, we describe a feature fusion method used to identify shops appearing in a photo.

Data Collection

The images illustrate a view of the storefront, including the style of decoration, tone of color and optionally a shop logo. We collected shop photos of 51 different shop classes from Google. These shop classes include a wide range of brands.

Retrieving Style Features

The AlexNet takes a patch with ﬁxed size 224\*224 as the input and our collected images usually have sizes greater than that, so we randomly pick 16 patches in an image and calculate each patch’s feature vector. In order to retrieve style features from the image, we ﬁne tuned the AlexNet using our data set to adapt it for our purpose. A 4096-dimension vector is hereby retrieved to represent style feature of the image. After that, we gain the ﬁnal style feature vector by taking the maximum of each bin.

Retrieving Text Features

To train the classiﬁer, we use 10000 ngram bins plus 4 geometrical features: w, h, w + h and w/h, which corresponds to the width, height, scale and shape of the box. This classiﬁer can distinguish true boxes from false ones with an accuracy of 93.89%. Finally, the candidate classiﬁed as true positive ones are used to generate text feature vectors.

Combining Styles and Text Features

One way is to get class score from ngram classiﬁer and style classiﬁer separately and add the two scores together as the ﬁnal class score.

Localization and Navigation

For the second shop name, operation above is repeated, but only node that is adjacent or equal to the nodes added to list in the last step is considered as candidates. Because the node-shop map is a many-to-many association, it requires at least two different shop observations to locate the user in most cases. For the ﬁrst shop name, nodes on AA are searched and nodes with landmarks corresponding to the shop name are added to candidate location list. That is to say, the system prompts the user to walk towards a certain sequence of shops that lies on the path leading to the destination. Inside a shopping mall, a user usually knows nothing about absolute directions (i.e., north, south, west and east) and there is no way to tell left and right simply from vision, so we make the navigation ”shop-based”. For path planning, the system gets user’s text input for destination and search for the node related to the destination. Dijkstra’s algorithm is then used to ﬁnd the shortest path between the destination node and the current node.

Conclusion

However, as there’s no large-scale storefront dataset publically available, we are not able to compare our method with others’ thoroughly at present. We showed by experiments that feature fusion can improve the accuracy of shop recognition. We put forward an automatic way of interpreting shopping mall ﬂoor plans and shop lists. This method utilizes both text feature extracted as ngram vectors and style feature extracted from a ﬁnetuned AlexNet. In this article, we proposed an indoor positioning system that works in shopping malls. We plan to continue this work by building an application system and test it in real shopping circumstances. While we see a great performance improvement by introducing feature fusion method, there are still some cases where feature fusion failed. The main problem with the current method is that both ngram feature and style feature are not robust enough for disambiguate a shop class.