ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

A Survey on Landmark Recognition with Deep Learning

Sabir A. Kazi[1], Kshitij A. Bulkunde[1], Abhik Chakraborty[1], Kajal P. Dhumal[1], Dr. Kishor Wagh[2]

B. E Student, Dept. of Computer Engineering, GHRCEM, Wagholi, Pune, India[1]

Head of Dept.(HOD), Dept. of Computer Engineering, GHRCEM., Wagholi, Pune, India[2]

ABSTRACT: An AR system will be used to process and identify a particular landmark's information by image processing using deep learning algorithm. The scanned image will be compared with sets of images created using grouping concept. Recognizing landmarks in sequences of images is a challenging problem for a number of reasons. Firstly, the view of any landmark varies according to the position of user from different angles. Finally, it is typically difficult to make use of accurate 3D information in landmark recognition applications.The alternative is to use image-based techniques in which landmarks are represented by collections of images which capture the "typical" appearance of the Landmark. The information most relevant to recognition is extracted from the collection of sub-images and used as the model for recognition. This process is called "visual learning." This image-based approach automatically learns the most salient landmarks in complex environments. It delivers a robust performance under a wide range of lighting and imaging angle variations.

KEYWORDS: Deep learning; Augmented reality; content matching; MQTT protocol; Android application; Grouping; SQLite database.

I. INTRODUCTION

The concept of Image recognition has been used frequently in various applications. Till date, various attempts are made to improve the functionalities of typical image based recognition techniques. Available researches show that with the help of machine learning, these techniques can improve the knowledge bases remarkably. Any typical recognition system uses a single image data for processing purpose which provides a moderate probability of content matching which in turn reduces the overall efficiency of the system. On the other hand, if pictorial data of any angle for specific landmarks is collected, the probability of perfect matches increases outstandingly. The threshold value decreases drastically as the dataset used will contain large metadata. This ultimately decreases user efforts and error generations.

Deep learning plays a very vital role for real-time applications. The efficiency of any deep learning algorithm depends on its training data. The more useable data available, more is the output availability. Also, compatibility for various devices is an important factor. When matching algorithms are used, main point to be considered is the processing time and space consumed for it. If pixel matching is discussed, it goes through the entire image and creates a sized histogram of pixel data. On the other hand, Content matching also creates a sized metadata set but it effectively reduces the processing time as it categorizes the image objects and uses these objects for comparison.

Since the era of internet, HTTP protocol has been the backbone for networking client-server models. This protocol was mandatory for and communication purposes. In later years, other protocols were developed to overcome the inefficiency of HTTP to transfer over wireless networks. For example, TCP/IP which controls the entire wireless data transmission. One of such is the MQTT (Message Queue Telemetry Transport) protocol. HTTP has a fixed packet size and transfers at a very slow rate in traffic. The ports that work for it are also standardized so any client will access only specific port number for transmission. Hence in overload cases, websites crash or transmit at minimum available rate. So MQTT was developed to try overcoming these issues. MQTT was developed by IBM developers. It provides high speed transmission for large scale data and is highly preferred for real-time transmission. Also, wide range of ports are available for connections so any specific port can be set.

To summarize, all above concepts explained i.e. image processing, content matching, MQTT etc. play a vital role in Deep learning and data transfer to output remarkable efficiency. The references mentioned here will give a quick overview of some of these concepts and a basic idea of how exactly things are expected to work. The proposed system focuses on giving and overview of an ideal system with expected maximum efficiency. Further, the actual working shows the complete picture and the literal flow of the system.

II. STATE OF ART

This paper of [1] Content and Context information Fusion (2011), Author[1]: Tao Chen, illustrates the use of mobile appliction for recognizing different landmarks using SVM algorithm . It recognizes the picture using contetnt matching . Information is extracted from the picture using Bow(bag of words) this content is matched with the image and GPS location and a result is generated . According to the survey made SIFT(scale invariant feature transform) is the most promising content analysis approach . However the large patches of SIFT in a single image makes computational cost of the image very high. To reduce this computational cost Bow has been used . However Bow has two major disadvantage (i)it ingnores the spatial layout information of the landmark image (ii)It does not discriminate the different foreground in the image. In order to eliminate the first problem a spatial pyramid matching scheme is used . It patitions the images into fine cells and then calculates the histogram inside cell of the image. However this pyramid does not discriminate between diiferent foreground. First the spatial pyramid decomposition is adopted to partition the image into a set of multi-resolution pyramid cells. The histogram of each cell is calculated using previously constructed codebook. A code word discriminationlearning approach is proposed to estimate the differentcode word representative capability for different categories.This information is then integrated with each cell's histogramto estimate the weighting value of each pyramid cell, whichindicates the different significance of foreground and background cells at each pyramid level. Further the weightingvalue of each cell is multiplied with its original BoWhistogram to generate a weighted histogram, and all theweighted histograms in an image are concatenated into a single descriptor as the image representation. Finally, SVM isused to train these descriptors into a classifier

This paper of [2] Survey on Mobile Landmark Recognition for Information Retrieval, Author[1]: Tao Chen, shows the use of mobile application for recognizing different landmark and to retrieve information from the picture. For classification purpose the is classified into different feature having global classification, local classification. The commonly used feature in mobile landmark recognition is global feature. Global feature differentiate picture according to its colour and texture. It uses multi-dimensional statistical responses of specific image filter to represent high dimensionality composed histogram feature.

It also uses SIFT for retrieving information from the picture. This feature is used in almost all the landmark recognition app Augmented technology is also used to display information from the picture. SIFT improves the accuracy and improves the computational time. Colour-edge histogram patch (CEHP) are used to select an optimal number of candidates in the first stage.

Probabilistic model-based feature uses probabilistic model to extract region information from the landmark. the commonly used generated models include evidence-based, latent dirichlet allocation, theme model based and entropy based. Probabilistic descriptor mainly focusses on the probability distribution of salient region in the images. Patches-based feature has been used to achieve good performance. SVM is used under discriminative based classification. An SVM classifier is trained for each category of landmarks using one-vs-all strategy.

In this paper, [3] Landmark Recognition Using Machine Learning, Author[1]: Andrew Crudge, input data of 193 images are given from google database. this data act as a training data set for the SVM. The training data set given to the SVM must contain vector that contain all the information of the images. To make each example the same feature dimension, the image must be cropped to an aspect of 5:2 ratio. To extract the feature from the image a HOG descriptor is used. HOG analyses the gradient orientation of the image. The images is divided into small cells. HOG descriptor also possesses several other useful properties for object detection. Drawback of HOG tool is that it does not adjust to the invariant in orientation of the images. SVM is used as machine learning classifier for this application. SVM was chosen as it works efficiently for high dimensional images. Drawback of SVM is it does not work on large number of datasets. Four classifier were compared by running a dataset of 100 images. But SVM attained a accuracy of 80%-90%. To detect a target building hidden within a larger image, the image is cropped into multiple overlapping cells with identical aspect ratios. Overall, the results are very encouraging, and they demonstrate that landmarks can be accurately identified from an image using a basic classification algorithm. An accuracy as high as 90% is attainable using a relatively small sample size

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

This paper is [4] Evaluation of Image-Based Landmark Recognition Techniques, Author[1]: Yutaka Takeuchi. Recognizing landmarks in sequences of images is a challenging problem for a number of reasons. First of all, the appearance of any given landmark varies substantially from one observation to the next. In addition to variations due to different aspects, illumination changes, external clutter, and changing geometry of the imaging devices are other factors affecting the variability of the observed landmarks. Finally, it is typically difficult to make use of accurate 3D information in landmark recognition applications. For those reasons, it is not possible to use many of the object recognition techniques based on strong geometric models. The color values are resampled using a standard equalization. Specifically, the histogram of color values is divided into eight classes of roughly equal numbers of pixels. The color image is then coded on eight levels using those classes Edges are computed using Deriche's edge detector, and edge elements are linked and expanded into line segments. The input image is a 160x120 normalized red image. Five histogram features are computed from the edge image. This registration problem can be made tractable through a few domain assumptions. First of all, the interest is in landmarks that are far from the observer. Second, let's assume that the images come from video sequences taken from a moving vehicle looking forward or from a transverse observer.

III. PROPOSED SYSTEM

The proposed system is to create a Landmark detection system with deep learning that will help common people to recognize any landmark like Hotels, School, Colleges, Hospital, etc.

This system will work on Android devices. It functions in such way that any technical and non-technical person can use this system. The system has three phases: (a) Login, (b) Capture the landmark image and recognize the image, (c) display the search result with its all information.

This system will use a client-server model. Client will be the user side which will use real-time data from the data server. For the communication between client and server, MQTT (Message Queue Telemetry Transport) protocol is used. It is "lightweight" protocol. It is specially designed for connection for remote location.

The input image from the user will be processed and stored in a categorized manner. For which, the concept of content matching will be preferred. Processing data pixel-by-pixel becomes a time as well as space consuming. So instead, it is stored in a tabular form which stores different category of objects that are present in any particular image. Hence, instead of dealing with typical pixel data, user's image can be recognized by matching its contents to the category tables of all images in the database.Also, to overcome the challenge of recognizing the landmark from any angle, "Grouping" concept come in handy. Here, instead of a single front side image, a set of images from all angles will be stored as a group for a specific landmark. Although it might produce a considerable size of tabular data, it will be way more accurate and efficient.

Finally, if the user is currently not close to the place of which they seek information, a search-by-text option will be given where user's input will be directly searched by the table names for all places in the database.

IV. WORKING OF SYSTEM

The application first asks the user for login so that the system can differentiate between user and client. the project has two applications viz. one for user and other for client. The client application will be privileged to update the data related to their institution which would be displayed along with the landmark. Clients include any specific institutions like hotels, workspaces, hospitals, monuments, etc. The user application will be used to take input data i.e. pictures or names and to provide respective description as output. Servers, along with the knowledge base will carry out content matching and formatting procedures. The input image will be compared with the dataset of sub-images of all locations to find a group with maximum similar categories and content match percentage.

The information database will be virtually divided by categories of institutions. As multiple institutions of similar categories generate similar type of data and structure, it will be much easier to target a particular type of data. Here, for content matching, a uniqueness or a specific object for any landmark will be stored specially. As this unique object differentiates the respective landmarks from one another, it almost ensures a perfect match or a higher probability of quick results. Still, a perfect match is next to impossible unless both compared images are the same. So, a threshold value is set so that maximum accuracy can be achieved while recognizing a particular image. This threshold value is the percent of minimum number of matched objects needed to declare any result as true.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

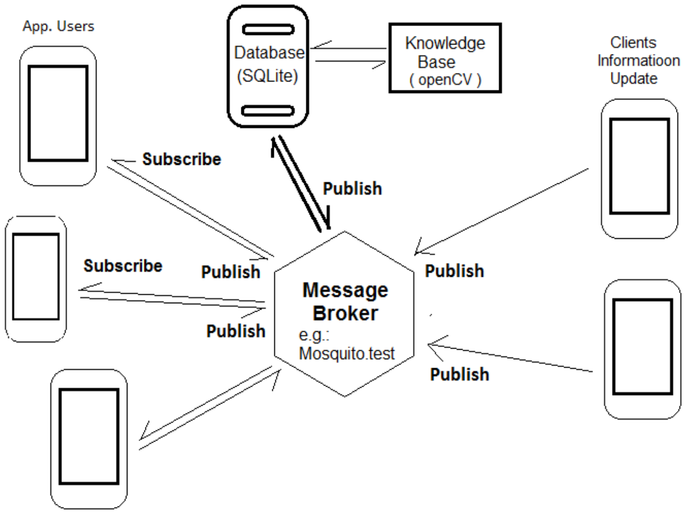

Fig.1.Architecture

When picture is clicked, it is stored temporarily in the mobile phone till the time it is not transferred to the server section.Transferring of the data from server to client and user is done by using MQTT protocol. This protocol is useful in machine to machine message transfer(M2M) and also in IoT. It works mainly on publish-subscribe method.publish–subscribe is a messaging pattern where senders of messages, called publishers, do not program the messages to be sent directly to specific receivers, called subscribers, but instead characterize published messages into classes without knowledge of which subscribers. Similarly, subscribers express interest in one or more classes and only receive messages that are of interest, without knowledge of which publishers.

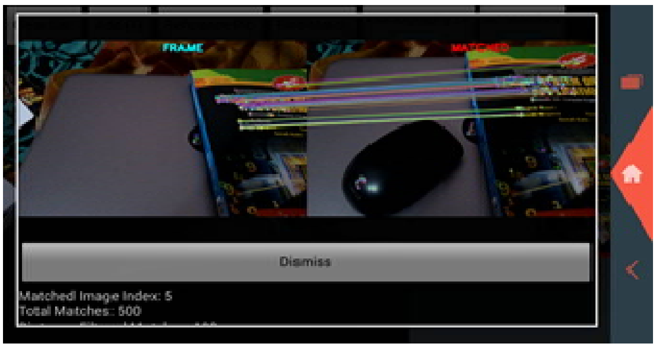

Fig.2. Content Matching between two images

Here, the user will publish a message in form of image data that is to be verified to all the servers randomly with a class or topic related to that image. Here, a program module called "Message Broker" is used. The main function of the broker is:

a. to coordinate the messages between all senders and receivers.
b. To translate message from the formal sender's protocol to the required receiver's protocol.
c. To invoke web services and handle errors.

Then the server containing or interested in the similar topic will subscribe to that message and establish a connection.

MQTT is used as it works efficiently on 2g/3g cellular network. Random port number can be set in MQTT which helps in reducing traffic on a single port. And thus, data does not have roadblock in processing and transfer of data. The client will be updating his data according to the available information. Database can be constructed using SQLite

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

because communication between same database can be easier. As SQL stores data in structured manner it will be easier for the system to process data.

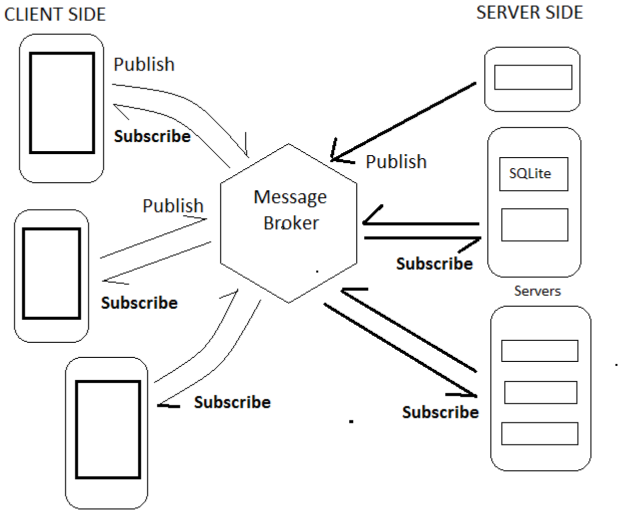

Fig.3.Working of MQTT

UML Diagrams:

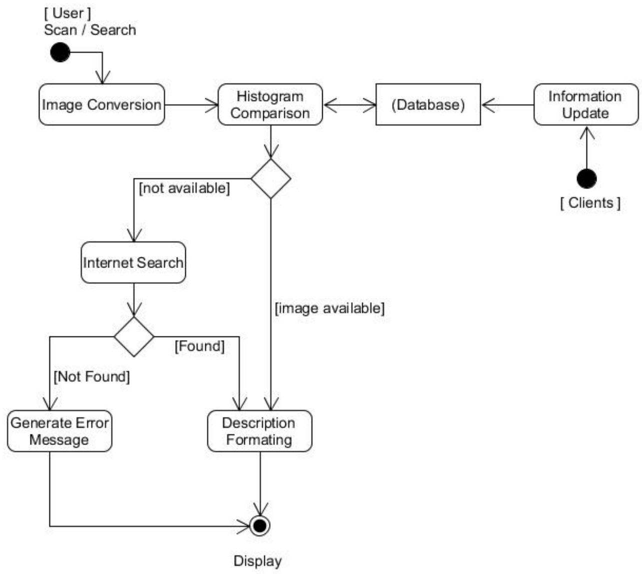

Fig. 4. Activity Diagram

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

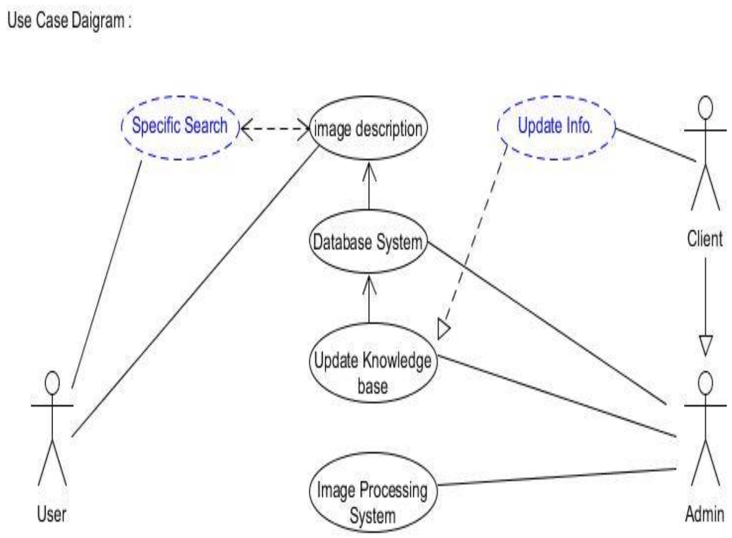

Fig. 5. Use Case Diagram

V. CONCLUSION AND FUTURE WORK

In the future, other feature extraction methods can be looked at that may give better accuracy but require fewer dimensions. An algorithm will also need to be developed that can automatically search and obtain data from a database. The project can ultimately become part of the back-end code for a feature-recognition smartphone app.

Further, there also lies the issue of storage management as this technique will trade with huge amount of dataset for maximum efficiency. This can be dealt by using or inventing appropriate algorithms to compress or manipulate stored data logically or physically such that it can be reframed on demand with reduced latency.

Also, a major affecting factor for the same is the transmission of data across the network. MQTT over TCP/IP is found much more effective at present for quick transfer of messages even on a giant scale but further researches can be made for attaining maximum flexibility and boost the transactions for better real time O-T-G experience.

Overall, the results are very encouraging, and they demonstrate that landmarks can be accurately identified from an image using a basic classification algorithm. An accuracy as high as 90% is attainable using a relatively small sample size. By using deep learning algorithm, work on large number of data set can be simplified. By using uniqueness in the picture the time required by the app to recognize any picture will be less. This will also help app to attain high accuracy in terms of final result. This app can be used in future in term of augmented reality to know the shape and structure of a particular building or monument. Even it can be used to recognize object which are used in our daily life.

REFERENCES

1. Tao Chen, Kim-Hui Yap, Lap-Pui Chau "Content and Context information Fusion for Mobile Landmark Recognition" IEEE pp. 978-1-4577-0031-6/11, 2011.
2. Tao Chen, Kui Wu, Kim-Hui Yap, Zhen Li, and Flora S. Tsai "A Survey on Mobile Landmark Recognition for Information Retrieval" IEEE pp. 978-0-7695-3650-7/09, 2009.
3. Yutaka Takeuchi, Martial Hebert "Evaluation of Image-Based Landmark Recognition Techniques" Carnegie Mellon University, 2009.
4. Tom Duckett, UlrichNehmzow "Performance Comparison of Landmark Recognition Systems for Navigating Mobile Robots" America Association for Artificial Intelligence, 2000.
5. David M. Chen "Memory-Efficient Image Databases for Mobile Visual Search" March 2014.
6. Aditya Srinivas Timmaraju and Anirban Chatterjee "Monulens: Real-time mobile-based Landmark Recognition" Stanford University Stanford,CA – 94305.
7. Filippo Galli "Landmark Recognition with Deep Learning" NEUROSCIENTIFIC SYSTEM THEORY" TechnischeUniversit at Munchen, 2016.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

8. Yan-Tao Zheng, Ming Zhao, Yang Song, Hartwig Adam, Ulrich Buddemeier, Alessandro Bissacco, Fernando Brucher, Tat-Seng Chua, and Hartmut Neven "Tour the World: building a web-scale landmark recognition engine"
9. Yunpeng Li David J. Crandall Daniel P. Huttenlocher"Landmark Classification in Large-scale Image Collections" Cornell University Ithaca, NY 14853 USA
10. A. M. Kaneko and K. Yamamoto, "Landmark recognition based on image characterization by segmentation points for autonomous driving," 2016 SICE International Symposium on Control Systems (ISCS), Nagoya, 2016, pp. 1-8.
11. J. Cao, T. Chen and J. Fan, "Fast online learning algorithm for landmark recognition based on BoW framework," 2014 9th IEEE Conference on Industrial Electronics and Applications, Hangzhou, 2014, pp. 1163-1168.
12. J. Cao et al., "Landmark recognition via sparse representation," 2015 IEEE International Conference on Digital Signal Processing (DSP), Singapore, 2015, pp. 1030-1034.
13. T. Guan, L. Cao, L. Cai and R. Ji, "Interactive on-device Mobile Landmark Recognition with compact binary codes," 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane, QLD, 2015, pp. 1141-1145.