

Vision Based Geo Navigation Information Retrieval

Asif Khan

School of Computer Science and
Engineering, UESTC
Chengdu 611731, China

Jian-Ping Li

School of Computer Science and
Engineering, UESTC
Chengdu 611731, China

Riaz Ahmed Shaikh

School of Computer Science and
Engineering, UESTC
Chengdu 611731, China

Abstract—In order to derive the three-dimensional camera position from the monocular camera vision, a geo-reference database is needed. Floor plan is a ubiquitous geo-reference database that every building refers to it during construction and facility maintenance. Comparing with other popular geo-reference database such as geo-tagged photos, the generation, update and maintenance of floor plan database does not require costly and time consuming survey tasks. In vision based methods, the camera needs special attention. In contrast to other sensors, vision sensors typically yield vast information that needs complex strategies to permit use in real-time and on computationally con-strained platforms. This research work show that map-based visual odometer strategy derived from a state-of-the-art structure-from-motion framework is particularly suitable for locally stable, pose controlled flight. Issues concerning drifts and robustness are analyzed and discussed with respect to the original framework. Additionally, various usage of localization algorithm in view of vision has been proposed here. Though, a noteworthy downside with vision-based algorithms is the absence of robustness. The greater parts of the methodologies are delicate to scene varieties (like season or environment changes) because of the way that they utilize the Sum of Squared Differences (SSD). To stop that, we utilize the Mutual Information which is exceptionally vigorous toward global and local scene varieties. On the other hand, dense methodologies are frequently identified with drift drawbacks. Here, attempt to take care of this issue by utilizing geo-referenced pictures. The algorithm of localization has been executed and experimental results are available. Vision sensors possess the potential to extract information about the surrounding environment and determine the locations of features or points of interest. Having mapped out landmarks in an unknown environment, subsequent observations by the vision sensor can in turn be used to resolve position and orientation while continuing to map out new features. In addition, the experimental results of the proposed model also suggest a plausibility proof for feed forward models of delineate recognition in GEO-location.

Keywords—*Vision; Geo-Navigation; Information Retrieval*

I. INTRODUCTION

These days, cell phones, e.g., cellular telephones and individual computerized associates are generally furnished with inserted GPS (Global Positioning Systems) and camera chips. Therefore, visual information connected with land or area labels can be effectively created in our day by day lives. Furthermore, numerous social sight and sound sharing stages, for example, Picasa and Flickr datasets empower clients to tag the areas of their transferred interactive media content on the land map. Furthermore, it is currently possible to derive the certain area labels from the sight and sound metadata like area names, phone numbers, and postal divisions. Such express and

understood geo labeling have brought a wealthy association among the sound substance, virtual sight and our physical world, bringing about an enormous yet steadily expanding measure of "geo-social interactive media" accessible on the web.

This geosocial interactive media has started another door for the scanning, hunt, examination and social mixed media mining, which includes a new "measurement" above the first visual features and data about data, means metadata connection. In the late decade, broad endeavors are given to the figuring, acknowledgment, mining, and investigation of geo-social interactive media, with developing applications going from area acknowledgment to scene rundown, tourism proposal, three dimensional building demonstrating, and city route and so forth. As we trust, the time has desired a convenient overview paper outlining late propel, central issues, and also open inquiries in geo-social sight and sound. In General, social mixed media alludes to the method for collaboration among individuals in which they make, offer, and/or trade mixed media data, e.g. pictures, recordings, sound, writings in virtual groups and organizes. We quickly audit the information source of social mixed media as beneath:

Explicit GPS tagging: A predominant extent of geo-social interactive media is delivered amid the visual substance creation strategy, e.g., amid the photograph catching by utilizing the inserted GPS chips alongside the cameras. Fig. 1 demonstrates some run of the mill gadgets furnished with implanted camera and GPS chip. An extensive asset originates from adjusting the time stamps between GPS gadgets and video streams [1][2].

The implicit geographical tagging: A new significant extent of geo-social media is delivered from the verifiable metadata depiction of sight and sound, from which the geological areas can be parsed utilizing the "area extraction" strategies [3][4][5][6]. Such verifiable portrayals might incorporate, yet not limit to, the URL address prefixes, dialect sorts, and land related things (such as, nation and city names, phone and post numbers). Case in point, picture metadata with "mass of china" demonstrates this picture is taken in Beijing, China.

Community Consensus: Web client group is the significant patron of geo-social sight and sound, which varies from the customary information acquisitions as far as its constantly expanding versatility and uncontrolled quality. From this point of view, the geo-social interactive media additionally mirrors the client inclination or "knowledge". One illustration is that the geo-labeled photographs in Panoramio or Flickr generally display an appropriation predisposition towards famous historic points. Inside of a given topographical area, the visual

insights additionally show a decent agreement to the agent points of interest, and in addition its delegate sees, which is normally the center in shooting for social interactive media clients. In said situations, by breaking down the metadata or visual measurements, the ubiquity of given points of interest or perspectives can be evaluated, that can be further abused with the end goal of touristic suggestion.



Fig. 1. GPS integrated mobile devices

II. LITERATURE REVIEW

Amartya *et al.* [7] emphasize on the design of an unmanned aerial system for handling of post disaster identification, effectively. Image segmentation is the key process in the domain of image processing. Rough set based image segmentation approach has been discussed by Roy *et al.* [8], it provides satisfactory method for the segmentation of image. Rough set theory is an approach to deal with vagueness and uncertainty. Chowdhuri *et al.* [9] presented a review for Rough set based Ad Hoc network. Images are used as data in the area of data mining. Current research work in the field of image mining and various techniques of this field has also been summarized by Dey *et al.* [10]. Watermarking is a security technique which provides data authentication. Acharjee *et al.* [11] proposed an algorithm to integrate the watermarking inside the motion vector. A unique algorithm has been discussed by Sakib *et al.* [12] for solving and exploring any kind of maze. Proposed system of maze mapping was based on coordinate system. For the extraction of shortest path and time Dijkstra's algorithm has been used, in adjacency list representation method whole maze as a graph can be saved. A prototype was developed in Thirumurugan *et al.* [13], of line following robot. Sensor based motors used to line following robot to keep track the line path.

Andersen *et al.* [14] explored the maturity of robot which is used for vacuum cleaning. Its described that vacuum cleaning robot was performed better cleaning as compared to conventional vacuuming. An integrated lane following system based on vision for a robot was presented by Huang and Houshang [15]. LBPE and Hough Transform used for lane position. Experimental results show that lane detection information is very much useful to navigate the mobile robot successfully. Coordinating multi robot system is still a problem in the field of robots. Montano and Suarez [16] proposed a solution to said problem. Proposed decoupled method can coordinate the multi participant robots in on-line mode. Two

robots were used to perform different tests in real environment and graphical simulation to assess the proposed approach.

In Kanayama [17] system for robot task design and management was developed. This system executes all processes from teaching to execution of task. An approach based on vision was proposed by Ismail *et al.* [18] for a mobile robot which follows the line. As a sensor pc camera was used, and via image segregation method image buffers was processed to output compulsory knowledge for the robot's controller. Experimental results shows that robot successfully follow and detect the provided path. Inchoative integration has been done for content based image retrieval by Khan *et al.* [19]. The proposed method just required very cheap hardware like webcam and PC. Based on the related algorithms and proposed approach software system was developed that shows efficiency and effectiveness of the system. Shukla and Tiwari [20] incorporated a approach based on fuzzy logic in the design of an autonomous mobile robot controller for steering and speed control. Portability and transportability are the advantages of proposed system.

A vision based scheme which comprises on two levels was presented for moving a mobile robot to catch a running target by Freda and Oriolo [21]. Nada *et al.* [22] presented a project Teleoperated Autonomous Vehicle. The main purpose of autonomous vehicle was to navigate a predefined route without touching any of the obstacles the vehicle may encounter. A machine vision based line follower robot was provided by Roy *et al.* [23]. A camera was used to image acquisition; algorithms and suitable image processing software were used to process the acquired image, robot was tracked on the basis of generated results. Genetic Algorithm (GA) based system was investigated by Wagner and Hagra [24] to change the type-2 membership functions parameter of interval type-2 Fuzzy Logic Controller purpose was robot control in real environment. By Shaikh *et al.* HMM can use for the content factors segmentation with content based image retrieval in real world [25].

III. APPLICATIONS SCENARIOS

There are expanding measure of various applications identified with geo-social mixed media, which can be quickly arranged into the accompanying four gatherings, i.e., (1) recognition of geographical location, (2) land mark mining and summarization, (3) Three dimensional city navigation and scene modeling.

A. The Geographical Location Recognition

To perceiving the topographical area of the info visual information, for example, a picture or a video arrangement. The acknowledgment is regularly accomplished by close copy visual inquiry from the question to the reference picture or video dataset. Vision-based area acknowledgment can be embraced as an integral or option answer for the GPS-based area acknowledgment. The acknowledgment is accomplished by means of coordinating the inquiry to an arrangement of reference photographs that are connected with land areas. This question might originate from a photograph or a video succession; either caught utilizing the cell phone or transferred by the web client. What's more, the comparable photographs of the question are recognized by utilizing close copy visual hunt

procedures, taking into account that location of the most comparative photo(s) can be distinguished as the acknowledgment yield. Databases that are used in this section are Zubud (Zurich city building dataset), Oxford buildings dataset, SCity contains more than 20,000 photos of street view, PKUBench (Dataset of Peking university).

B. Landmark Mining and Recommendation

Intends to finding delegate historic points from geo-labeled photographs. The mined points of interest can be compressed and prescribed for visual city tourism or online. For this situation, the topographical labels can be either unequivocal or certain. What's more, the mining is principally taking into account investigating the agreement among the connection and substance of such photographs. Currently, there is an expanding enthusiasm for acknowledgment and mining of points of interest from geo-tagged photographs from photograph sharing sites, for example, Flickr and Picasa [26,27,28,29,30]. As a kind of client produced content, these photographs commonly uncover the conduct agreement of online networking clients, bringing about an overwhelming part of milestones in these photographs. From the point of view of the visual appearance, such client agreement additionally compares to the close copy measurements of photographs labeled close-by. Mining milestones from social mixed media represents a wide assortment of utilizations, going from city scene rundown and three dimensional demonstrating to tourism suggestion. Geo-tagged Flickr photos in 20 worldwide capital cities including Boston, Beijing, Chicago, London, New York, Rome etc., Blog photos is a dataset covers social media blogs, Online landmark corpus enables users to upload photos to build three dimensional landmark models online.

C. Three Dimensional Scene Modeling and City Navigation

It plans to remake the three dimensional models, for individual building or for the whole metropolitan city. The prime thought is to do picture correspondence to assess the profundity or three dimensional structures from pictures inside of a sure area, which are recognized utilizing their GPS or area labeling. The vast majority of the world's well known areas have been captured under different conditions, both starting from the earliest stage from the air. It has delivered an extremely rich symbolism of historic points and urban areas empowering three dimensional three dimensional demonstrating and route. Given the three dimensional models of a sure scene or historic point, it turns out to be much simpler for vision-based expanded reality. The increase originates from taking out the past equivocalness in coordinating genuine scenes to two dimensional images. Beneath we survey related work on three dimensional demonstrating and expanded (virtual) reality route by utilizing geo-social sight and sound. City-scale modeling and Landmark-scale modeling are used as geographical modeling.

IV. PROBLEM DOMAIN

In vision based methods, the camera needs special attention. In contrast to other sensors, vision sensors typically yield vast information that needs complex strategies to permit use in real-time and on computationally constrained platforms. Vision is certainly not the only sense nature is using for successful navigation and localization. Some animals, bats for

example, developed sonar-like sensors with remarkable properties. Most accessible for our understanding are daily situations when we merge visual cues with vestibular cues such as accelerations and angular velocities. When looking downwards from a high altitude, we are not able to determine the depth of the scene because of our narrow stereo baseline. Only when we start to move sufficiently we provoke sufficient visual stimulus for depth perception. This phenomenon is also known as Structure from Motion (SfM) in Computer Vision.

V. METHODOLOGY

In geo image mining it is very likely that shadow will be mistaken for objects, since all objects are donated objects.

So geo mining also required to consider for to create the possibility clearly show object for that in this reason work consider two parameter which resolve shadow condition i.e. Brightness and density. Fig. 2 demonstrated geo navigation.

Where

$$\text{Brightness} = \frac{R + G + B + NIR}{4} \tag{1}$$

And

$$\text{Density} = \frac{\sqrt{\#P_r}}{H\sqrt{\sigma_x^2 + \sigma_y^2}} \tag{2}$$

So diameter of square object $\sqrt{\#P_r}$ with diameter of ellp

$$\sqrt{\sigma_x^2 + \sigma_y^2} \tag{3}$$

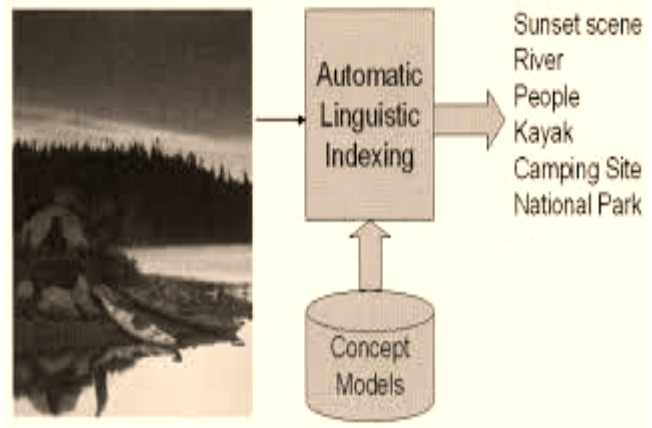


Fig. 2. Geo navigation model

So fuzzy rule used to detect clarity of object in geo map are as

$$\mu_B(x) = \begin{cases} 1 & \text{for } X < M \\ 1 - \frac{x-m}{3\sigma} & 0 \leq x - m \leq 3\sigma \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

$$\mu_D(x) = \begin{cases} 1 & \text{for } x < 1 \\ 1 - \frac{x-1}{.2} & \text{for } 0 \leq x-1 \leq 0.2 \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

So as per above fuzzy function using to detect area, where medium fuzzy function use to detect building detection, low level fuzzy function used to detect vehicle detection. Geo process show in Fig. 3.

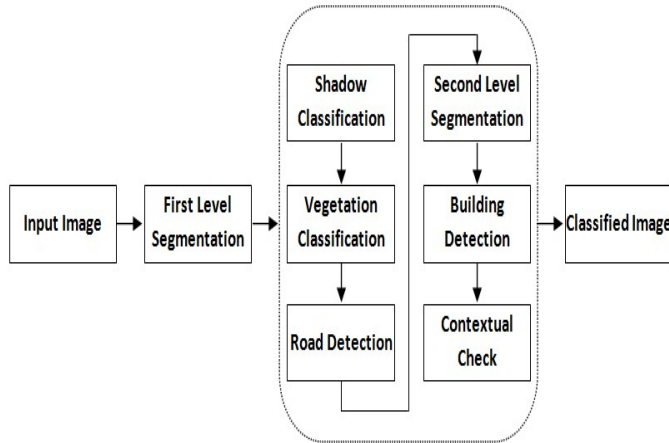


Fig. 3. Geo process

Results and discussions

The whole dataset has been physically annotated to create a complete ground truth for 11 unique landmarks. Each contains 5 queries, 55 queries in total to assess the accuracy of location recognition. One of the accompanying marks are relegated to every picture, i.e.: (1). Excellent: A decent, clear photo of the object; (2). Good: 25% and more of the article is unmistakably visible; (3). Poor: The item is not present; (4). Trash: Less than 25% of the article is visible, or there are elevated amounts of distortion and occlusion. Fig.4 shows the examples of landmark photos from Geo-tagged Flickr photos.



Fig. 4. Flickr photo collection

VI. CONCLUSION

This research work demonstrates that map-based visual odometer strategy derived from a state-of-the-art structure-from-motion framework is particularly suitable for locally stable, pose controlled flight. Issues concerning drifts and

robustness are analyzed and discussed with respect to the original framework. Also this paper proposed a multiple usage localization algorithm based on vision only. One of the principle main impetuses for geological mindful social interactive media is the success of client group, which to a great extent diminishes the broad expense of human work in delivering or gathering the media content and metadata. Such information scale brings another inquiry. In common information scale, because of the information and predisposition, existing methodologies typically embraced parametric forecast models. In any case, whether parametric models are so suitable for displaying the enormous and continually developing information scale is flawed. In such situations, there is more prospect on researching the un-parametric models.

ACKNOWLEDGMENT

This paper was supported by the National Natural Science Foundation of China (Grant No.61370073), the National High Technology Research and Development Program of China (Grant No.2007AA01Z423).

REFERENCES

- [1] Ji R., Xie X., Yao H. and Ma W.-Y., Mining city landmarks from blogs by graph modeling, ACM Multimedia, 2009.
- [2] Jia M., Fan X., Xie X., Li M. and Ma W.-Y., Photo-to-search: Using camera phones to inquire of the surrounding world, Mobile Data Management, 2006.
- [3] Buyukkokten O., Cho J., Garcia-Molina H., Gravano L. and Shivakumar N., Exploiting geographic location information of web pages, ACM SIGMOD Workshop on the Web and Databases, 1999
- [4] Li H., Srihari R. K., Niu C. and Li W., Location normalization for information extraction, COLING, 2002.
- [5] Li H., Srihari R. K., Niu C. and Li W., InfoXtract location normalizations: A hybrid approach to geographic references in information extraction, International Workshop on the Analysis of Geographic References, 2003.
- [6] Wang C., Xie X., Wang L., Lu Y. and Ma W.-Y., Detecting geographic locations from web resources, ACM Geographical Information Systems Workshop, 2005.
- [7] Amartya Mukherjee, Sayan Chakraborty, Ahmad Taher Azar, Soumya Kanti Bhattacharyay, Basukinath Chatterjee, Nilanjan Dey (2014) Unmanned Aerial System for Post Disaster Identification. The International Conference on Circuits, Communication, Control and Computing (I4C), 21 - 22 November 2014, MSRIT, Bangalore, India, pp. 247 – 252.
- [8] Roy P., Goswami S., Chakraborty S, Azar AT, Dey N (2014) Image Segmentation Using Rough Set Theory: A Review. International Journal of Rough Sets and Data Analysis, 1(2): 62-74.
- [9] Chowdhuri S, Roy P, Goswami S, Azar AT, Dey N (2014) Rough Set Based Ad Hoc Network: A Review. International Journal of Service Science, Management, Engineering, and Technology, 5(4): 66--76.
- [10] Dey N, Karaa WBA, Chakraborty S, Banerjee S, Salem MAM, Azar AT (2015) Image Mining Framework and Techniques: A Review. International Journal of Image Mining, Inderscience, 1(1): 45-64.
- [11] Acharjee S, Chakraborty S, Samanta S, Azar AT, Dey N, Hassanien AE (2014) Highly secured multilayered motion vector watermarking. In: AE Hassanien, M Tolba, AT Azar (eds.) Advanced Machine Learning Technologies and Applications: Second International Conference, AMLTA 2014, Cairo, Egypt, November 28-30, 2014. Proceedings, Communications in Computer and Information Science, Vol. 488, Springer-Verlag GmbH Berlin/Heidelberg.
- [12] Sakib, S., Chowdhury, A., Ahmed, S., T., Hasan, S., I., "Maze Solving Algorithm for Line Following Robot and Derivation of Linear Path Distance from Nonlinear Path", Int'l Conference on Computer and Information Technology, pp. 478-483, 2014.

- [13] Thirumurugan, J., Kartheeswaran, G., Vinoth, M., Vishwanathan, M., "Line Following Robot for Library Inventory Management System", International Conference on Emerging Trends in Robotics and Communication Technologies, pp. 1-3, 2005.
- [14] Andersen, K., N., Medaglia, R., Gimpel, G., Sjolín, P., Mikkelsen, L., S., "Robots on the Move from the Production Line to the Service Sector: The Grand Challenges for Contractors, Workers, and Management", eChallenges 2-2010 Conference Proceedings, pp. 1-7, 2010.
- [15] Huang, X., Houshang, N., "A Vision Based Autonomous Lane Following System for a Mobile Robot", IEEE International Conference on Systems, Man and Cybernetics, pp. 2344-2349, October 2009.
- [16] Montano, A., Suarez, R., "An On-Line Coordination Algorithm for Multi-Robot Systems", IEEE Conference on Emerging Technologies and Factory Automation, pp. 1-7, September 2013.
- [17] Kanayama, K., Mizukawa, M., Iwaki, S., Matsuo, S., "A Robot Task Design and Management System for Industrial Applications", International Conference on Advanced Robotics, pp. 687-692, July 1997.
- [18] Ismail, A., H., Ramli, H., R., Ahmad, M., H., Marhaban, M., H., "Vision Based System for Line Following Mobile Robot", Symposium on Industrial Electronics and Applications, pp. 642-645, 2009.
- [19] Khan, A.; Jian-Ping Li; Shaikh, R.A.; Deep, S.; Memon, M.H.; Kumar, K., "Inchoative integration of content based image retrieval: Shodhani", 11th International Computer Conference Wavelet Active Media Technology and Information Processing, pp.289-292, Dec. 2014.
- [20] Shukla, S., Tiwari, M., "Fuzzy Logic of Speed and Steering Control System for Three Dimensional Line Following of an Autonomous Vehicle", International Journal of Computer Science and Information Security, Vol. 7, No. 3, pp. 101-108, March 2010.
- [21] Freda, L., Oriolo, G., "Vision Based Interception of a Moving Target with a Nonholonomic Mobile Robot", Robotics and Autonomous Systems Vol. 55, pp. 419-432, 2007.
- [22] Nada, E., S., Abd-Allah, M., Ahmed, A., Tantawy, M., "Teleoperated Autonomous Vehicle", International Journal of Engineering Research and Technology, Vol. 3, Issue 7, pp. 1088-1095, July 2014.
- [23] Roy, A., I., Raja, B., K., Chakkaravathy, G., Prakash, C., A., "Line Following Robot Based on Vision Techniques", international Journal of Advanced Technology in Engineering and Science, Vol. 3, Issue No. 2, pp. 266-269, February 2015.
- [24] Wagner, C., Hagras, H., "A Genetic Algorithm Based Architecture for Evolving Type-2 Fuzzy Logic Controllers for Real World Autonomous Mobile Robots", IEEE International Conference on Fuzzy Systems, pp. 1-6, July 2007.
- [25] Shaikh, R.A.; Deep, S.; Jian-Ping Li; Memon, M.H.; Khan, A.; Kumar, K., "Content factors segmentation with CBIR in real world", 11th International Computer Conference Wavelet Active Media Technology and Information Processing, pp.297-300, Dec. 2014.
- [26] Kennedy L., Naaman M. and Ahern S., How Flickr helps us make sense of the world: context and content in community contributed media collections, ACM Multimedia, 2007
- [27] Hays J. and Efros A., IMG2GPS: Estimating geographic information from a single image, CVPR, 2008
- [28] Ji R., Xie X., Yao H. and Ma W.-Y., Mining city landmarks from blogs by graph modeling, ACM Multimedia, 2009
- [29] Zheng Y.-T., Zhao M., Song Y. and Adam H., Tour the world: Building a web-scale landmark recognition engine, CVPR, 2009
- [30] Li X., Wu C., Zach C., Lazebnik S. and Frahm J.-M., Modeling and recognition of landmark image collections using iconic scene graphs, ECCV, 2008.