DOI: 10.2352/ISSN.2470-1173.2016.7MOBMU-299

Turbo Fusion of LPQ and HOG Feature Sets for Indoor Positioning Using Smartphone Camera

Jichao Jiao IEEE Member, Zhongliang Deng IEEE Member, Jun Mo, Cheng Li
school of Electronic Engineering, Beijing University of Posts and Telecommunications
Beijing, China

Abstract

More recently, the smartphone intergrated powerful camera is an efficient platform for location-wareness. The matching of smartphone recordings with a database of geo-referenced images allows for meter accurate infrastructure-free localization. However, for high accuracy indoor positioning using a smartphone, there are two constraints that includes: (1) limited computational and memory resources of smartphone; (2) user's moving in large buildings. These constraints are also typically more severe for systems that should be wearable and used indoors. To address these issues, we proppose a novel smartphone camera-based algorithm for supporting a scalability and high accuracy indoor positiong service. In order to obtain an accurate image matching, we proppose a new feature descriptor that efficiently fused of HOG and LPQ feature. The novel feature is the local phase quantization of a salient HOG visualuizing image. The specific properties of this feature is robust in the indoor scenarios. In order to reduce the network latency and communications traffic, we introduce a basestation based indoor positiioning system for providing a coarse location. Comparing to other states of art methods, experimental results show that our algorithm allowed instantaneous camera-based indoor positioning with very low requirements on the available network connection.

INTRODUCTION

Indoor positioning is considered an enabler for a variety of applications, such as guidance of passengers on airports, conference attendees, visitors in shopping malls, and for many novel context-aware services, which can play a signicant role for monetarization. The demand for an indoor positioning service or indoor LBS (iLBS) has also accelerated given that people spend the majority of their time indoors [2]. Over the last decade, researchers have studied many indoor positioning techniques [18]. In addition, with the development of the integrated circuit technology, multi-sensors, for example, camera, Earths magnetic field, WiFi, Bluetooth, inertial module, have been integrated in smartphones. Therefore, smartphones are powerful platforms for location-awareness.

The traditionally used outdoor localization method **G**lobal **N**avigation **S**atellite **S**ystem (GNSS) is not available in indoor environments,even though navigation tasks on street level are very precise. A catalog of alternative localization techniques has been investigated, such as infrared- [13], sensor- [15][4], wireless- [36][7], communication basestation-based technologies[34] , Pseudolite [25] or visual markers [11]. However, most those technologies, however, relying on wireless technology, faces issues in the presence of RF interference (RFI), and interference of Non Line of Sight (NLOS) caused by dense forests, urban canyons, terrain [17]. Moreover, some of those technologies work in a limited area like inertial-sensor based approaches, or some need a particular environmental infrastructure and augmentation like Locata that is a pseudolite positioning system. Therefore, smartphone camera-based indoor positioning is a promising approach for accurate indoor positioning without the need for expensive infrastructure like acess points or beacons.

The key method of camera-based localization is image mtaching. Images taken by a smartphone camera are matched to previously acquired reference images with known position and orientation.The matching of smartphone recordings with a database of geo-referenced images allows for meter accurate infrastructure-free [28]. According to the matched reference image, the location of the smartphone is calculated. In mobile indoor scenarios that are shown by Fig.3, the users usually walk during positioning and navigation procedure. Therefore, the captured images by smartohone cameras are scaled, rotated, even blured because of hands shaking. Moreover, Recently, most of researchers focus on invariant features extraction. Ravi and his co-workers extracted color histograms, wavelet decomposition and image shape for image matching to locate the user's position [26]. Kim and Jun proposed a method based on image colorfu histogram feature for positioning by using augmented reality tool[12]. However, the positioning accuracy of those two methods would work inefficiently in the varying light and crowed scenarios. In order to extract the invariment features, SIFT and its improved algorithms are widely used for image-based indoor locatilization. Kawaji et al. used PCA-SIFT feature for railway museum indoor positioning. Werner and his colleague proposed a camera-based indoor positioning by using SURF feature for speeding up the image matching [32]. Li and Wang [16]introduced A-SIFT feature for image matching achieved by RANSAC, which increased the matching accuracy. Tian and his co-workers [28] proposed a similar method to [16] for indoor positioning, However, those two complex computational methods is not suitable for smartphone-based indoor positioning. This is beause of limited computational resources of mobile devices. Zachariah and Jansson [33] extracted the edge-based features from the visual tags image, and those features are fused with inertial information for indoor navigation. Kazemipur et al. [11] used the Sobel filter intergrating mean structural similarity index for estimating the arrival of angle and height during the indoor localization. However, those two methods need additional visul marks for assisting smartphone camera for detecting features, which increases the indoor positioning cost. Meanwhile, all of those research work mainly focus on improving image matching accuracy. Some of these algorithms are, however,

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.7MOBMU-299

quite demanding in terms of their computational complexity and therefore not suited to run on mobile devices, which need smartphones with high hardware configuration. Aathough smartphpone phones are inexpensive, they have even more limited performance than the aforementioned Tablet PCs. Phones are embedded systems with severe limitations in both the computational facilities and memory bandwidth. Therefore, natural feature extraction and matching on phones has largely been considered prohibitive and has not been successfully demonstrated to date [29]. To address these issues, Opdenbosch et al. [28] used the improved Vector of Locally Aggregated Descriptors (VLAD) image signature and emerging binary feature descriptor BRIEF to achieve the smartphone camera-based indoor positioning. Besides, in order to reduce the overall computational complexity, they proposed a scalable streaming approach for loading the reference images onto the phones. Different with their method, this paper prosposed a efficient feature descriptor named Turbo Fusing Histograms of Oriented Gradients (HOG) and Local Phase Quantization (LPQ) Salient feature (TFHLS). The TFHLS features are extracted from the partial image which are salient image regions, and they are invarient to the illumination, scale, rotation and blur caused by cmera shaking. Moreover, a wireless-based indoor positioning method TC-OFDM are introduced to calculate the coarse positions for supporting the floor number to the smartphone, which would reduce the number of images downloaded onto the smartphones. By using this approach, our camera-based indoor positioning algorithm results in the reduction in computational complexity, hardware requriment, and network latency.

This paper is organized as follows to achieve our investigations. First of all, we discuss the reated work on HOG and LPQ feature extraction in Section . Then, we introduce our image feature extraction based on fusing HOG and LPQ in Section. After that, we test the proposed algorithm on the TUM Indoor Dataset [34] and BUPT Indoor Dataset collected by our lab, and the evalution of our algorithm is also shown in this section. Finally, in Section we conclude the paper and provide an future work on possible extensions.

RELATED WORK

Finding efficient and discriminative descriptors is crucial for indoor complex scenarios. HOG descriptor was proposed by Dalal .et al. for human detection [1]. The main idea behind HOG is based on the local edge information [6]. Because of its efficient performance, HOG feature are widely used in human detection [23] [35], face recognition [14] [31], and image searching [27]. All of those applications show that HOG feture is invariant to the illumination. According to our experiment, HOG feature is not robust when the humans are crowded and the images are blurred. Wang and his co-workers combined the HOG and Local Binary Pattern (LBP) features for human detection [30]. However, they calculated that their detector cannot handle the articulated deformation of people.

Recently, LPQ is insensitive to image blurring, and it has proven to be a very efficient descriptor in face recognition from blurred as well as sharp images [5] [24][6]. LPQ was originally designed by Ojansivu and Heikkila similarly to the LBP methodology as a texture descriptor[22]. In our opinion, robust and efficient image matching requires several different kinds of appearance information to be taken into account, suggesting the use of heterogeneous feature sets. In our prposed algorithm, the HOG features are extracted from the salient regions, and LPQ featrues are extracted from the HOG vVisualizing image. Therefore, the HOG and LPQ are intergated for buiding a efficient feature that is TFHLS for indoor image matching.

PROPOSED SMARTPHONE CAMERA-BASED INDOOR POSITIONING

The smartphone camera-based indoor positioning procedure by using TFHLS feature is shown in fig.1. Meanwhile, the framework of our smartphone camear-based indoor positioning system is shown by fig.2.

Flowchart.png

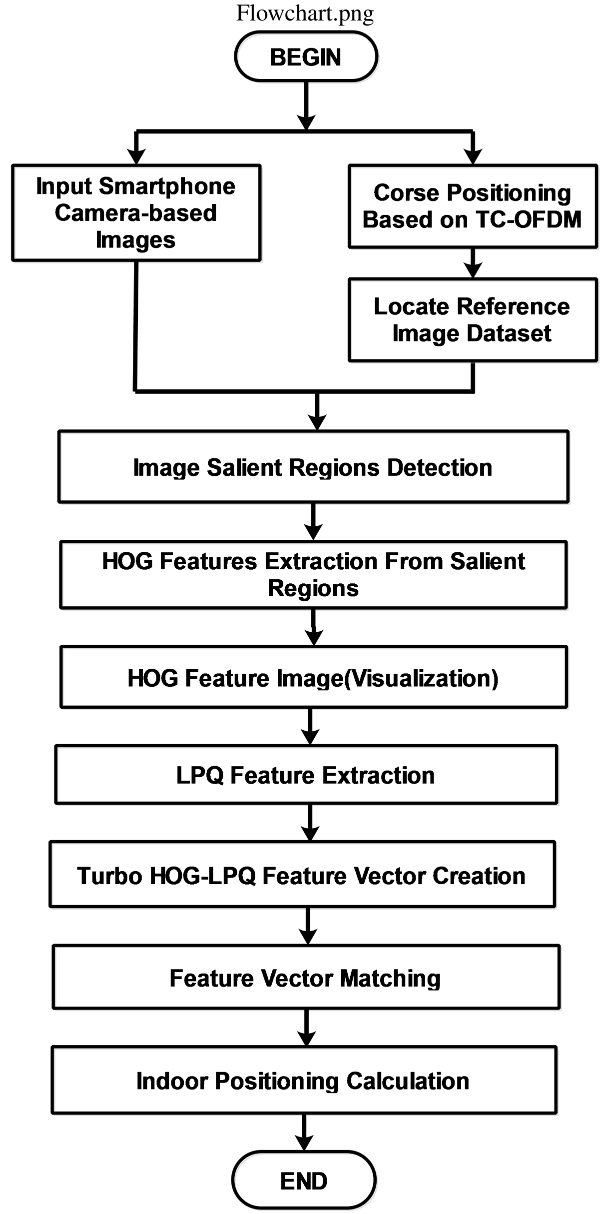


Figure 1. The module of smartphone camera-based indoor positioning.

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.7MOBMU-299

and Position Module.png

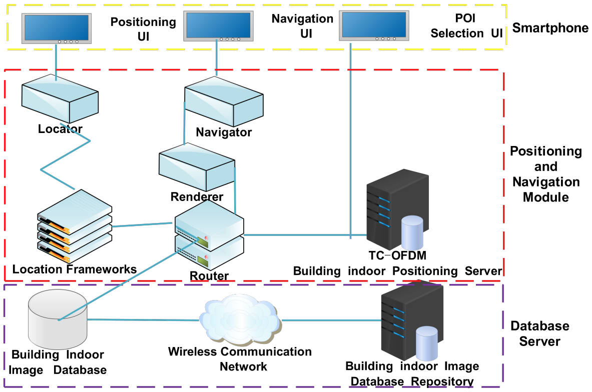


Figure 2. The framework of smartphone camera-based indoor positioning.

Study Materials

In our indoor positioning scenario, the reference database contains images captured inside buildings and can be queried with an image taken by a smartphone camera. The position information attached to the most similar reference view serves as the location estimate. It is noticed that in order to test and evaluate the proposed algorrthm, two databsets are used. The frist one is supported by Technische Universität München (TMU) [28]. The researchers in TMU emploied the virtual view approach proposed by Huitl et al. [8] for building a meter-accurate localization system with the possibility and calculating viewing angles. In TMU dataset, there are 54, 896 reference views, which covers 3, 431 positions with 1 meter accuracy. The examples of TMU dataset are shown by fig.3. Another dataset is collected by our lab who caputred 1000 indoor images using smartphone cameras in BUPT campus. The examples of BUPT dataset are shown by fig.4. Different with TMU dataset in calculating the reference positions, a static measurement system based on TC-OFDM and Beidou Real Time Kinematic (RTK) is introduced. By using this system, the scalable locations with positioning accuracy $(0.6 \sim meter)$ are obtained. The BUPT dataset covers four buildings and results in total of 2,189 positions.

Turbo HOG-LPQ Feature Extraction Approach

HOG Features Extraction and Visulization

Compare to the original HOG, the integrated HOG feature without trilinear interpolation is easier and faster to be computed, which was improved in [9]. However, the HOG's performance would be worse. Therefore, we introduced a constrained trilinear interpolation approach to replace the general trilinear interpolation. A novety 5×5 convolution kernel that similar to [30] is built to be implemented. For a 8-bit image, the kernal template is shown by eq.1.

$$Conv_{HOG} = \frac{1}{256}\begin{bmatrix} 1 & 3 & 4 & 3 & 1 \\ 3 & 6 & 8 & 6 & 3 \\ 5 & 12 & 16 & 12 & 5 \\ 3 & 6 & 8 & 6 & 3 \\ 1 & 3 & 4 & 3 & 1 \end{bmatrix} \quad (1)$$

Moreover, in order to reduce the space complexity of the integral image method, the kernal in 1 is convoluted with the salient rectangle not the whole orignal image.

LPQ Feature Extraction From HOG Visulization Imgae

More recently, LPQ features were used for face recognition. LPQ was originally built by Ojansivu and his co-workers as a texture descriptor, which was similar to the Local Binary Pattern (LBP) [20]. While, LPQ is robust to image blurring, and it has proven to be a very efficient descriptor in face recognition from blurred as well as sharp images[19][3]. Moreover, according to our previous research, the LPQ features extracted from indoor scences images are less precise in positioning than used in face recognition. For the indoor images with low textures, it is difficult to extract salient feature because of low contrast between objects and background. Meanwhile, the contrast of the HOG integral image is good to be used for LPQ extraction. In this paper, we intrduced a HOG visualizing method proposed by Vondrick al et.[21]. Different with their complex method, an simplified method based on eq.2 is prposed, which remain the performance as the orignal visualizing method in [21].

$$\phi^{-1}(y) = \underset{x \in \mathbb{R}^D}{argmin} \parallel \phi(x) - y \parallel^2 \quad (2)$$

where $x \in \mathbb{R}^D$ is an salient rectangle subimage and $y = \phi x$ is the corresponding HOG feature descriptor. In this paper, HOG feature visualization is posed to be a feature inversion procedure. In order to optimizing 2, we used gradient-descent strategies by numerically evaluating the derivative in image space with Least Squares method. After inverting HOG features into an image Y_HOG, LPQ features are extracted from Y_HOG by using a simple scalar quantizer **??**. LPQ feature is based on quantifying the Fourier transform phase by considering the sign of each component in Fourier coefficients $G(x)$.

$$q_i(x) = \begin{cases} 1 & if \quad g_i(x) \leq 0 \\ 0 & \quad otherwise \end{cases} \quad (3)$$

where $g_i(x)$ is the i_{th} component of $G(x)$. Then the phase information of the 8-bit HOG visualizing image described using 4.

$$f_{LPQ}(x) = \sum_{n=1}^{8} q_n 2^{n-1} \quad (4)$$

The final LPQ features are used as a feature vector to represent an indoor subimage.

Human Positioning by Using TFHLS Feature Matching

The main advantage of the binarization, apart from a reduced memory footprint, is a very fast matching process using the Hamming distance $h(x,y)$. This distance can be computed very efficiently using intrinsic processor instructions, which are also available on modern mobile devices. For the ranking process, we match every subsection b_i of the rectangle subimage separately. The final score f for query image q and database image d is calculated as:

$$f(b_q, b_d) = \frac{1}{\sqrt{|F_q||F_d|}} \sum_{i \in F_1 \cap F_d} 1 - \frac{h(b_{q,i}, b_{d,i})}{l'} \quad (5)$$

DOI: 10.2352/ISSN.2470-1173.2016.7MOBMU-299

(a) Low textures (b) High textures (c) Blurred Image (d) Building Hall (e) Hallway (f) Illumination change

Figure 3. *Exemplary queries for all classes from TMU.*

(a) Low textures (b) High textures (c) Blurred Image (d) Building Hall (e) Hallway (f) Illumination change

Figure 4. *Exemplary queries for all classes from BUPT.*

where F_q and F_b are the sets of visual words observed in the database and query image, respectively. When dealing with larger datasets, e.g., whole cities, we employ a TC-OFDM indoor positioning system to locale the building where user is, which provides fast approximate nearest neighbor search.

EXPERIMENTAL RESULT

Query Dataset and Setup Description

We recorded a query set of 128 images captured by a iPhone 6 smartphone with manually annotated position information. The images are approximately 5 megapixels in size and are taken using the default settings of the iPhone 6 camera application. Furthermore, the images consist of landscape photos either taken head-on in front of a store or at a slanted angle of approximately 30 degrees. After obtaining the images, Next, we run the remaining query images with successful retrieved database images through the pose estimation part of the pipeline. In order to characterize pose estimation accuracy, we first manually ground truth the position and pose of each query image taken. This is done by using the CAD map of the buildings in BUPT and distance measurements recorded during the query dataset collection. For a detailed evaluation, the query set has been split into classes that is the same with the TMU databse: high texture, low texture, hallways, ambiguous objects and building structure, where each query can be assigned to more than one class (fig.3 and fig.4). It is should be known that we ignore the orientation information calculation.

Our method was implemented by using Matlab 2015a, and this method was coded by integrating *C#* and matlab. It is noticed that the camera-based positioning method proposed by [28] is used to compare with our proposed method, and the test data and matlab code of that method are both supported by Opdenbosch.

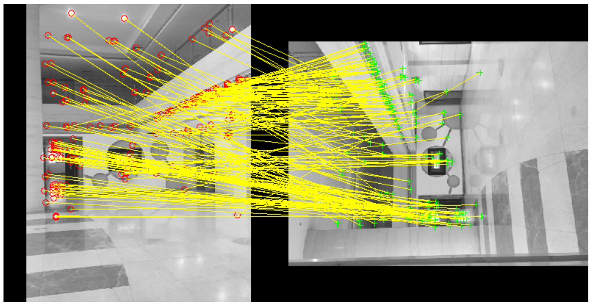
Feature Matching Evaluation

In order to identify optimal parameters for the approach described above, several experiments are conducted with varying settings. fig.5 summarizes the performace of comparing the TFHLS feature matching to the method proposed by [28] . In this experiment, we successfully match 113 of 128 images to achieve a retrieval rat of 93%. As shown in fig.5 (a), successful retrival usually involves matching of object textures in both query and database imgaes. According to fig.5 (b), we can find that our proposed TFHLS feature is efficient to match the blurred images. As shown in Table1, the proposed method achieves to match the images of TMU databse with a highest success in 15.7*ms* for each image. where LS means linear search, LSH means locality-sensitive hashing. In order to measure the running time of feature

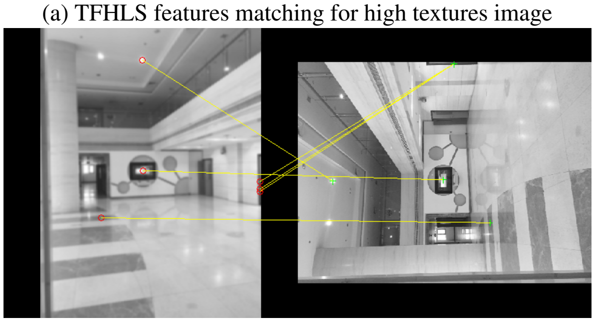
Matching Result

Setp	Running Time	Matching rate
TFHLS Detector	89%	13.2 *ms*
FAST Detector	68%	0.98 *ms*
FAST Detector	68%	0.98 *ms*
BRIEF Descriptor	73%	4.77 *ms*
SURF Detector	82%	232.6 *ms*
BVLAD Matching(LSH)	85%	53.74 *ms*
BVLAD Matching (LS)	87%	100.17 *ms*

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.7MOBMU-299



(a) TFHLS features matching for high textures image



(b) TFHLS features matching for blurred image

Figure 5. *TFHLS features matching for BUPT images.*

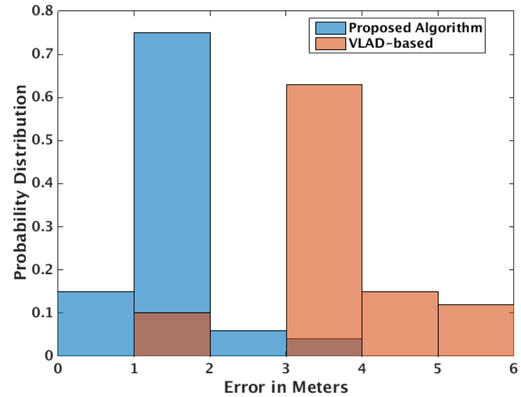
extraction, a laptop intergrated an Intel Core i7 x64 system with 2.8GHz is used.

Positioning Result Evaluation

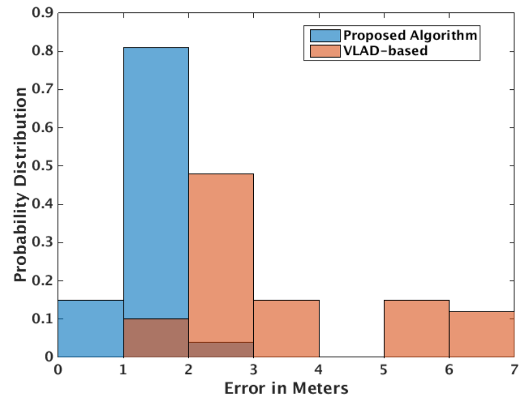
fig.6 summarizes the performance of the location information estimation, and the comparison result, which was tested on TMU database, between our method and the BVLAD-based method is shown in fig.6(a), and the positioning result tested on BUPT database is shown by fig.6(b). From fig.6(a) and fig.6(b), we are able to localize the position to within sub-meter level of accuracy for over 56% of the query images. Furthermore, 85% of the query images are successfully localized to within two meters of the ground truth position. As seen in fig.5 (a), when the location error is less than 1 meter, the TFHLS features of corresponding corridor signs present in both query and database images are matched together well. Conversely, in less accurate cases of pose estimation where the location error exceeds 4 meters, more false matching corresponding features between query and database images. Moreover, we find that the TFHLS detector extracted more features than [28] even throgh the images are blurred, which is shown in fig.6(b). As shown in fig.7, we plot the estimated and ground truth locations of the query images onto the New Research Buildings 2D floorplan. As seen from this figure, there is close agreement between the two. The Root Mean Square Error (RMSE) between estimated and ground-truth positioning results is 1.253 meters.

CALCULATION

We present a scalable and effeceient mobile camera-based localization system. To this end, we propose a modified version feature of combining HOG and LPQ descriptors, which is based on texture and phase features and jointly addresses the problem



(a) Positiong result based on TMU dataset



(b) Positiong result based on BUPT dataset

Figure 6. *Positioning performance comparison.*

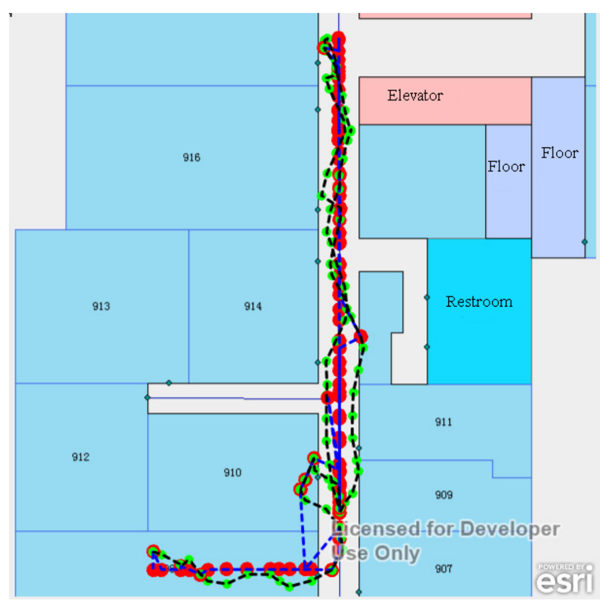


Figure 7. *The module of smartphone camera-based indoor positioning.*

DOI: 10.2352/ISSN.2470-1173.2016.7MOBMU-299

of limited computational capacity, as well as the required memory footprint. For rapid and accurate matching, we extract the features from the salient sub-images, which reduces the featture searching space. Those are also our main contribution. Moreover, in order to provide an efficient approach of fetching the reference images from the database server, we employs TC-OFDM indoor positioning for supporting the corase positioning knowledge related camera location for the smartohone, where make the feature space in a certain radius.This results in a significant reduction in data rate down to 70% of the communication traffic, while maintaining the full positioning performance. According to the test on the BUPT databse, the Root Mean Square Error (RMSE) between estimated and ground-truth positioning results is 1.253 meters, which shows that our smartphone camera-based indoor positioning is precise and accuracy. The other contributions of this paper lead to an indoor localization system intergating camera and RF module of a smpartphone, which allows instantaneous camera-based indoor positioning with very low requirements on the available network connection.

ACKNOWLEDGEMENTS

The Project Sponsored by The National High Technology Research and Development Program of China (NO.2014AA123103, NO.2015AA124103) and National Natural Science Foundation of China (NO.61401040).

References

[1] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.

[2] Marta C Gonzalez, Cesar A Hidalgo, and Albert-Laszlo Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.

[3] Diego Gragnaniello, Giovanni Poggi, Carlo Sansone, and Luisa Verdoliva. Local contrast phase descriptor for fingerprint liveness detection. *Pattern Recognition*, 48(4):1050–1058, 2015.

[4] Dominik Gusenbauer, Carsten Isert, and Jens Krösche. Self-contained indoor positioning on off-the-shelf mobile devices. In *Indoor positioning and indoor navigation (IPIN), 2010 international conference on*, pages 1–9. IEEE, 2010.

[5] J. Heikkila, V. Ojansivu, and E. Rahtu. Improved blur insensitivity for decorrelated local phase quantization. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 818–821, Aug 2010.

[6] Janne Heikkilä, Esa Rahtu, and Ville Ojansivu. Local phase quantization for blur insensitive texture description. In *Local Binary Patterns: New Variants and Applications*, pages 49–84. Springer, 2014.

[7] Chung-Hao Huang, Lun-Hui Lee, Chian C Ho, Lang-Long Wu, and Zu-Hao Lai. Real-time rfid indoor positioning system based on kalman-filter drift removal and heron-bilateration location estimation. *Instrumentation and Measurement, IEEE Transactions on*, 64(3):728–739, 2015.

[8] Robert Huitl, Georg Schroth, Sebastian Hilsenbeck, Florian Schweiger, and Eckehard G. Steinbach. Virtual reference view generation for cbir-based visual pose estimation. In *Proceedings of the 20th ACM Multimedia Conference, MM '12, Nara, Japan, October 29 - November 02, 2012*, pages 993–996, 2012.

[9] Zhu, Qiang and Yeh, Mei Chen and Cheng, Kwang Ting and Avidan, Shai. Fast Human Detection Using a Cascade of Histograms of Oriented Gradients. In *Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1491-1498, 2006.

[10] Jichao Jiao, Zhongliang Deng, Baojun Zhao, John Femiani, and Xin Wang. A hybrid method for multi-sensor remote sensing image registration based on salience region. *Circuits, Systems, and Signal Processing*, 33(7):2293–2317, 2014.

[11] B. Kazemipur, Z. Syed, J. Georgy, and N. El-Sheimy. Vision-based context and height estimation for 3d indoor location. In *Position, Location and Navigation Symposium - PLANS 2014, 2014 IEEE/ION*, pages 1336–1342, May 2014.

[12] Jongbae Kim and Heesung Jun. Vision-based location positioning using augmented reality for indoor navigation. *Consumer Electronics, IEEE Transactions on*, 54(3):954–962, August 2008.

[13] Chunhan Lee, Yushin Chang, Gunhong Park, Jeaheon Ryu, Seung-Gweon Jeong, Seokhyun Park, Jae Whe Park, Hee Chang Lee, Keum-shik Hong, and Man Hyung Lee. Indoor positioning system based on incident angles of infrared emitters. In *Industrial Electronics Society, 2004. IECON 2004. 30th Annual Conference of IEEE*, volume 3, pages 2218–2222. IEEE, 2004.

[14] Joel Z Leibo, Qianli Liao, and Tomaso Poggio. Subtasks of unconstrained face recognition. In *International Joint Conference on Computer Vision, Imaging and Computer Graphics, VISAPP,(Lisbon, Portugal), SCITEPRESS*, 2014.

[15] Binghao Li, Thomas Gallagher, Andrew G Dempster, and Chris Rizos. How feasible is the use of magnetic field alone for indoor positioning? In *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on*, pages 1–9. IEEE, 2012.

[16] Xun Li and Jinling Wang. Image matching techniques for vision-based indoor navigation systems: performance analysis for 3d map based approach. In *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on*, pages 1–8, Nov 2012.

[17] Y Li and C Rizos. Triple integration of gps, locata and ins for seamless integrated navigation in urban environments [z]. In *Proceedings of 2010 International Symposium on Inertial Technology and Navigation, NanJing*. Citeseer, 2010.

[18] Hui Liu, Houshang Darabi, Pat Banerjee, and Jing Liu. Survey of wireless indoor positioning techniques and systems. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 37(6):1067–1080, 2007.

[19] Huu-Tuan Nguyen and A. Caplier. Patch based local phase quantization of monogenic components for face recognition. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 229–233, Oct 2014.

[20] V. Ojansivu and J. Heikkila. Blur invariant registration of rotated, scaled and shifted images. In *Signal Processing Conference, 2007 15th European*, pages 1755–1759, Sept

DOI: 10.2352/ISSN.2470-1173.2016.7MOBMU-299

2007.

[21] Vondrick, Carl and Khosla, Aditya and Malisiewicz, Tomasz and Torralba, Antonio. HOGgles: Visualizing Object Detection Features. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 1–8. IEEE, 2014.

[22] V. Ojansivu, E. Rahtu, and J. Heikkila. Rotation invariant local phase quantization for blur insensitive texture analysis. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4, Dec 2008.

[23] Yanwei Pang, Yuan Yuan, Xuelong Li, and Jing Pan. Efficient hog human detection. *Signal Processing*, 91(4):773–781, 2011.

[24] M. Pedone and J. Heikkila. Local phase quantization descriptors for blur robust and illumination invariant recognition of color textures. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 2476–2479, Nov 2012.

[25] Nonie Politi, L Yong, K Faisal, C Mazher, Jonas Bertsch, J Cheong, Andrew Dempster, and Chris Rizos. Locata: A new technology for high precision positioning. In *European Navigation Conference, Naples, Italy*, 2009.

[26] Nishkam Ravi, Pravin Shankar, Andrew Frankel, Ahmed Elgammal, and Liviu Iftode. Indoor localization using camera phones. In *Mobile Computing Systems and Applications, 2006. WMCSA'06. Proceedings. 7th IEEE Workshop on*, pages 49–49. IEEE, 2006.

[27] Babak Saleh, Mira Dontcheva, Aaron Hertzmann, and Zhicheng Liu. Learning style similarity for searching infographics. In *Proceedings of the 41st Graphics Interface Conference*, GI '15, pages 59–64, Toronto, Ont., Canada, Canada, 2015. Canadian Information Processing Society.

[28] D. Van Opdenbosch, G. Schroth, R. Huitl, S. Hilsenbeck, A. Garcea, and E. Steinbach. Camera-based indoor positioning using scalable streaming of compressed binary image signatures. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 2804–2808, Oct 2014.

[29] Daniel Wagner, Gerhard Reitmayr, Alessandro Mulloni, Tom Drummond, and D. Schmalstieg. Pose tracking from natural features on mobile phones. In *Mixed and Augmented Reality, 2008. ISMAR 2008. 7th IEEE/ACM International Symposium on*, pages 125–134, Sept 2008.

[30] Xiaoyu Wang, Tony X Han, and Shuicheng Yan. An hog-lbp human detector with partial occlusion handling. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 32–39. IEEE, 2009.

[31] Xin Wei, Gongde Guo, Hui Wang, and Huan Wan. A multi-scale method for hog-based face recognition. In *Intelligent Robotics and Applications*, pages 535–545. Springer, 2015.

[32] Martin Werner, Moritz Kessel, and Chadly Marouane. Indoor positioning using smartphone camera. In *Indoor Positioning and Indoor Navigation (IPIN), 2011 International Conference on*, pages 1–6. IEEE, 2011.

[33] D. Zachariah and M. Jansson. Fusing visual tags and inertial information for indoor navigation. In *Position Location and Navigation Symposium (PLANS), 2012 IEEE/ION*, pages 535–540, April 2012.

[34] Deng Zhongliang, Yu Yanpei, Yuan Xie, Wan Neng, and Yang Lei. Situation and development tendency of indoor positioning. *Communications, China*, 10(3):42–55, 2013.

[35] Qiang Zhu, Mei Chen Yeh, Kwang Ting Cheng, and Shai Avidan. Fast human detection using a cascade of histograms of oriented gradients. In *Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1491–1498, 2006.

[36] Han Zou, Hao Jiang, Xiaoxuan Lu, and Lihua Xie. An online sequential extreme learning machine approach to wifi based indoor positioning. In *Internet of Things (WF-IoT), 2014 IEEE World Forum on*, pages 111–116. IEEE, 2014.

Author Biography

Jichao Jiao, received the PhD degrees in signal and information processing from Beijing Institute of Technology (2013). He is an Assistant Professor in Beijing University of Posts and Telecommunications since 2013. His research activities include computer vision, GNSS navigation and indoor positioning.

Deng Zhongliang, received the MSc degree in manufacturing engineering from Beihang Universuty, and PhD degree in mechanical manufacture from Tsinghua University, China. He is a professor and doctoral supervisor at School of Electronic Engineering in Beijing University of Posts and Telecommunications now. He holds the position of director of research with the Laboratory of Intelligent Communication, Navigation and Micro/Nano-Systems (ICNMNS). He is also the executive vice President of BUPT. His research interests include indoor and outdoor seamless positioning, GNSS, satellite communications, MEMS and multimedia.