Analysis of Image Classification and Annotation Methods for Smartphone

Lekhraj D. Vilhekar[1], Dr. M. B. Chandak[2]

[1]Department of Information Technology, Government Polytechnic, Nagpur, MS, India

[2]Professor, Dept of CSE, Ramdeobaba College of Engg. & Management, Nagpur, MS, India

Abstract: *Rapid development in computing technology brings computers and humans to be seamlessly integrated in future. The innovation of smart phone has driven computing era towards ubiquitous and pervasive computing. Recognizing human activity has garnered a lot of interest and has raised significant researches- concerns in identifying contextual information useful to human activity recognition. The central elements of pervasive computing is, the ability to recognize and understand its users dynamically for which classification, object separation and annotation required. This paper discusses about classification algorithms used in smart phone-based human activity. Existing technologies pertaining to smart phone-based researches in human activity recognition will be highlighted and discussed.*

Keywords: Classification & Annotation algorithms, Human Activity Recognition (HAR), Smart phones

1. Introduction

In daily life, smart phone has embedded with built-in sensors that capable to sense contextual information of its users supported with wide range capability of network connections.

Digital pictures and videos are undergoing an explosion, especially with the proliferation of high-quality digital cameras embedded in mobile devices. As these pictures get stored in online content warehouses, the need to search and browse them is becoming crucial. Furthermore, the growing sophistication in textual search is raising the expectations from image retrieval—users are expecting to search for pictures as they do for textual content. Efforts to engage humans for labeling/tagging pictures may be a stop-gap solution, but is not likely to scale in the longer run. The whole volume of content is growing at dramatic speeds, and its complete dependence on a pair of human eyes and brain is likely to become the bottleneck.

Computers and humans have converged into one inseparable entity in daily life. Computers' existence has become ubiquitous in human's daily life. This factual reasoning has led into emerging computing perspective so called pervasive computing where computing process is already existing everywhere and every time in everyday objects. One of the central elements of pervasive computing is the ability to recognize and understand its users dynamically. That ability can be provided by obtaining the current and updated information about the user [1]. The recent information about users is basically the information about context of users. Context of users can be any information and status in regards to users' status, user's location, user's environment, and so on. Context-aware computing is then introduced to address the challenges to recognize and to understand them more especially under dynamic changes from users. There are growing interests on recognizing context generated from human notably known as human activity recognition.

2. Real World Failures

Nowadays, mobile phones are going to be the important resource for computing, personal sensing and communication. Hereby, it is an attempt to overcome the problem of automatic image annotation.

By involving smart phones, researchers aim towards successful human activity recognition with image classification and annotation. With their advantageous characteristics compared to wearable sensors, human activity recognition of particular users is expected to be performed unobtrusively when observing people behaviors, their activities and their surrounding environments. Additionally, being different from wearable sensor-based systems which require separated processing unit, by using smart phone it is expected to have context processing integrated into one device where data has been collected. This will promise faster and more reliable decision result as it no longer requires specific communication medium between processing unit and sensor. Response with particular services once recognition is successful also now being possible as smart phone has wide range of connectivity options. With those several possibilities, eventually, smart phone-based system will enable human activity recognition being portable even during intense users' mobility activities or movements.

TagSense is tagging the image using the sensors which are not that much reliable to correctly recognize the person or activity in the image as every person in the image must have the mobile with those sensors facility [1].

There are still many learning methods which are not coming with expected output that is actually required[2].The fields of image processing and face recognition have made significant progress but still automatic image annotation is long standing problem.

3. Issues to be Consider

One of the contexts that researchers want to recognize is users' activity by using image classification and annotation. Image classification and annotation results into Activity recognition (AR) is a more specific and highly demanded field in context-aware computing where researchers aim to recognize any actions from users. Typical activities to detect vary from mechanical process like activities of daily livings (ADL) until socio-spatial processes like meetings. The advances in activity recognition will enable computer systems to bring better services and user experiences towards their users such as in healthcare services and living services or other socio services. Activity recognition basically concerns about the users and/or their surrounding environment. In socio-spatial process like meetings, an activity recognition system must be able to address both users and their environment. In a facet of activity recognition, Human Activity Recognition (HAR) solely focuses on humans as users in its concern.

In practical, Image classification and annotation, activity recognition can be conducted aiming to detect single user activity recognition from recognizing ADL of a person up until multi-user activity recognition like in surveillance and monitoring situation.

More specifically, a group of researchers focus more on data fusion techniques for many types of sensors deployed. However, there are still researches conducted that focus on solely vision-based approaches using video/image-based system. Researchers use both real-time video feeding and stored video files as input to reveal contexts in human recognition. Researchers propose an idea that multi-view from multiple cameras fusion for whole area observation has yielded reasonable performance for human activity recognition in their smart studio laboratory.

4. Current Methodology in use-

TagSense
TagSense[1] , the different sensors are likely to capture the —moments" across multiple sensing dimensions. Laughing may be more naturally detectable via the microphone; dancing may exhibit an accelerometer signature; light sensors may easily discern between indoor and outdoor environments. TagSense [1] needs to 1) identify the individuals in the picture—since the person's(taking the picture) phone gathers sensed information from all phones within wireless range, it is unclear which of the phones were in the picture. 2) Sensor readings gathered from different phones need to be mined to identify activities and contextual information. 3) The energy budget for sensing, communicating, and computing needs to be optimized to facilitate wide-scale adoption.

TagSense adopts three mechanisms. 1) When people explicitly pose for the picture, TagSense extracts a pause signature from the accelerometer readings. This pause signature correlates well with the timing of the photo-click, and is found to be mostly absent in people who are not posing for the picture. 2) People in the picture are often faced toward the camera. TagSense leverages the phones' and camera's compass directions to infer a —mutually facing" relationship; this heuristic improves the confidence of the posing signatures. As will be evident later, unknown and time-varying phone orientations make the problem difficult. 3) For pictures in which the subjects do not pose explicitly, the TagSense camera takes multiple snapshots. The motion vectors for the subjects are computed from the sequence of snapshots, and then correlated with the motion derived from the phones' accelerometer/compass readings. Phones that exhibit a good correlation (between the visual and acceleration dimensions) are used for tagging.

TagSense requires the content in the pictures to have an electronic footprint that can be captured over at least one of the sensing dimensions. If the objects do not present such footprints, one has to rely on the visual domain alone for detecting them, which is the measure disadvantage for the TagSense.

Image Annotation Genetic Algorithm
Automatic image annotation (AIA)[3] , for labeling images accurately, more various features containing low-level image features, textual tags of images have been extracted so far; however, not whole features give useful information for each conception. Feature selection as one of the important preprocessing methods, which contain the optimization of feature descriptor weights and the selection of an optimum subset feature descriptor, are desirable to improve the performance of image annotation by decreasing the feature dimension properly. An automated annotation based method to solve AIA in three separate phases, which is named Image Annotation Genetic Algorithm (IAGA). Principally, GA as feature selection in the first phase to solve the high dimensions problem, in the next phase , Multi-Label KNN algorithm to weight neighbors and generate a novel weighted matrix, and in the third phase use GA to combine the results and assign the related words to new images.

Classification Algorithms

One of the important components in HAR is classification algorithm used to classify different activities and actions based on the user inputs. The algorithm usually is executed either on a workstation or user's smart phone. The selection of classification algorithm is based on the capability of the processing platform to execute the algorithm. Moreover, the evaluation method is used to measure the performance of the classification algorithm. Most researchers use supervised classification algorithms. The algorithms are trained with labeled samples to generate classification model. Then the model will be used for classification of input data. From the survey, the most popular algorithms are Decision Trees, k-Nearest Neighbor, Naïve Bayes, Support Vector Machine and Neural Network. Brent Longstaff et al. in [5] introduce algorithms using semi-supervised and active learning

methods. They investigated two semi-supervised learning methods which are self-learning and co-learning. Self-learning use one classifier to classify unlabeled data. If the confidence level of the prediction by the classifier is high, the data will be labeled with the prediction. Another semi-supervised learning method called co-learning that uses multiple classifiers to classify unlabeled data. Since supervised classification algorithms need intensive computation to generate models from training data, thus most of the activity based on the user's environment inputs. These kinds of algorithms are suitable to be implemented in smart phone because its need less computation resources. On the other hand, some researchers [8], [13] generate classification model by executing the algorithm at a workstation. Then bring the model into smart phones for

classification of input data. Evaluating the classification algorithms used for HAR is very important since it shows which algorithm performs better. From our survey, the popular evaluation methods are n-fold cross validation (commonly 10-fold), precision and recall measures, F-measures and accuracy. Claudia Nickel et al. in [9] used error rates from biometric matching, which are False Non-match Rate (FNMR) and False Match Rate (FMR) to evaluate their method. FNMR may be compared with false positive and FMR with false n.

5. Comparison of Classification Algorithms in Activity Recognition

Year	Ref	Classification Algorithms Evaluation Method Processing	Platform
2014	[1]	Multidimensional sensing heuristic correlate to accelerometer measurements	Phone
2014	[2]	MPEG compact descriptors for visual search	Phone
2011	[19]	Hidden Markov Models 4-fold cross validation and precision/recall measures	Phone
2011	[17]	Multiclass Logistic Regression 4-fold cross validation and precision/recall measures	Server
2011	[20]	Transfer learning Embedded Decision Tree 10 times 10-fold cross validation	*Server*
2011	[28]	Hidden Markov Chain	*Server*
2011	[18]	Decision Tree, Naïve Bayes, Random Forest, Logistics Regression, RBF Network, Support Vector Machine 10-fold cross validation	*Server*
2011	[30]	Smoothed Single-layer Hidden Markov Models F-measure	*Server*
2011	[14]	Decision Tree	*Phone*
2011	[16]	Hidden Markov Model false non match rate (FNMR), false match rate (FMR)	*Server*
2010	[23]	Support Vector Machine F-measure, precision and recall	*Server*
2010	[22]	Decision Tree, Logistic Regression and Multilayer Neural Networks 10-fold cross validation	*Server*
2010	[24]	Recurrent Fuzzy Inference Systems Accuracy	*Server*
2010	[25]	k-Nearest Neighbour, Direct Density, Class Local Outlier Factor, Local Classification Factor Precision and recall	*Phone*
2010	[29]	Artificial Neural Network Accuracy	*Server*
2010	[21]	Active Learning, Self-learning, and Co-learning 10-fold cross validation	*Server & phone*
2010	[27]	Decision Tree, Bayesian Network, Naïve Bayes, k-Nearest Neighbour, Support Vector Machine 10-fold cross-validation, mean and 95% confidence interval	*Server*
2010	[26]	Decision Tree, k-Nearest Neighbours and Sequential Minimal Optimization Accuracy	*Server*
2010	[13]	Gaussian Mixture Model (GMM) and Support Vector Machine Accuracy and Error	*Phone*

References

[1] Chuan Qin; Xuan Bao; Choudhury, R.R.; Nelakuditi, S., "TagSense: Leveraging Smartphones for Automatic Image Tagging," *Mobile Computing, IEEE Transactions on* , vol.13, no.1, pp.61,74, Jan. 2014

[2] Zhu Songhao; Li Xiangxaing; Li Zhuofan; Hu Juanjuan, "An improved approach for image annotation," *Control Conference (CCC), 2014 33rd Chinese* , vol., no., pp.4673,4678, 28-30 July 2014

[3] Richang Hong; Meng Wang; Yue Gao; Dacheng Tao; Xuelong Li; Xindong Wu, "Image Annotation by Multiple-Instance Learning With Discriminative Feature Mapping and Selection," *Cybernetics, IEEE Transactions on* , vol.44, no.5, pp.669,680, May 2014

[4] O. W. H. Wu, A. a T. Bui, M. a Batalin, L. K. Au, J. D. Binney, and W. J. Kaiser, ―MEDIC: medical embedded device for individualized care.,"

[5] T.-seong Kim, ―Multi-modal Sensor-based Human Activity Recognition for Smart Homes Author biography," no. February, pp. 157-174, 2011.

[6] C. Thurau and V. Hlavac, ―Pose primitive based human action recognition in videos or still images," 2008 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8, Jun. 2008.

[7] R. Poovandran, ―Human activity recognition for video surveillance," 2008 IEEE International Symposium on Circuits and Systems, pp. 2737-2740, May 2008.

[8] G. D. Abowd, C. G. Atkeson, J. Hong, S. Long, R. Kooper, and M. Pinkerton, ―Cyberguide : A mobile context-aware tour guide," vol. 3, pp. 421-433, 1997.

[9] K. Cheverst, N. Davies, K. Mitchell, A. Friday, and C. Efstratiou, ―Developing a Context-aware Electronic Tourist Guide: Some Issues and Experiences," vol. 2, no. 1, pp. 17-24, 2000.

[10] L. Bao and S. S. Intille, ―Activity Recognition from User-Annotated Acceleration Data,‖ Most, pp. 1-17, 2004.

[11] M. Fahriddin, M. G. Song, J. Y. Kim, and S. Y. Na, ―Human Activity Recognition Using New Multi-Sensor Module in Mobile Environment,‖ 2011.

[12] S.-I. Yang and S.-B. Cho, ―Recognizing human activities from accelerometer and physiological sensors,‖ 2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, pp. 100-105, Aug. 2008.

[13] E. Miluzzo, M. Papandrea, N. D. Lane, H. Lu, and A. T. Campbell, ―Pocket , Bag , Hand , etc . - Automatically Detecting Phone Context through Discovery.‖

[14] G. Bieber, A. Luthardt, C. Peter, and B. Urban, ―The Hearing Trousers Pocket – Activity Recognition by Alternative Sensors,‖ Science, 2011.

[15] S. C. Wangberg, E. Arsand, and N. Andersson, ―Diabetes education via mobile text messaging.,‖ Journal of telemedicine and telecare, vol. 12 Suppl 1, pp. 55-6, Jan. 2006.

[16] C. Nickel, C. Busch, and M. Möbius, ―Using Hidden Markov Models for Accelerometer-Based Biometric Gait Recognition,‖ Applied Sciences, pp. 58-63, 2011.

[17] D. Riboni and C. Bettini, ―COSAR: hybrid reasoning for context-aware activity recognition,‖ Personal and Ubiquitous Computing, vol. 15, no. 3, pp. 271-289, Aug. 2010.

[18] L. Sun, D. Zhang, and N. Li, ―Physical Activity Monitoring with Mobile Phones,‖ pp. 104-111, 2011.

[19] Y.-seol Lee and S.-bae Cho, ―Activity Recognition Using Hierarchical Hidden Markov Models on a Smartphone with 3D Accelerometer,‖ pp. 460-467, 2011.

[20] Z. Zhao, Y. Chen, J. Liu, Z. Shen, and M. Liu, ―Cross-People Mobile- Phone Based Activity Recognition,‖ Science, pp. 2545-2550, 2006.

[21] B. Longstaff, S. Reddy, and D. Estrin, ―Improving Activity Classification for Health Applications on Mobile Devices using Active and Semi-Supervised Learning,‖ Computer.

[22] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, ―Activity Recognition using Cell Phone Accelerometers,‖ Human Factors, vol. 12, no. 2, pp. 74-82, 2010.

[23] L. Sun, D. Zhang, B. Li, B. Guo, and S. Li, ―Activity Recognition on an Accelerometer Embedded Mobile Phone with Varying Positions and Orientations,‖ pp. 548-562, 2010.

[24] M. Berchtold, M. Budde, H. R. Schmidtke, and M. Beigl, ―AnExtensible Modular Recognition Concept That Makes Activity Recognition Practical,‖ pp. 400-409, 2010.

[25] N. Bicocchi, M. Mamei, and F. Zambonelli, ―Detecting activities from body-worn accelerometers via instance-based algorithms,‖ Pervasive and Mobile Computing, vol. 6, no. 4, pp. 482-495, Aug. 2010.

[26] S. L. Lau, I. König, K. David, B. Parandian, C. Carius-düssel, and M. Schultz, ―Supporting Patient Monitoring Using Activity Recognition with a Smartphone,‖ Immanuel, no. April, pp. 810-814, 2010.

[27] S. L. Lau and K. David, ―Movement Recognition using the Accelerometer in Smartphones,‖ Communication, pp. 1-9, 2010.

[28] E. Martin et al., ―Enhancing Context Awareness with Activity Recognition and Radio Fingerprinting,‖ pp. 263-266, 2011.

[29] A. M. Khan, Y. Lee, and S. Y. Lee, ―Human Activity Recognition via An Accelerometer-Enabled-Smartphone Using Kernel Discriminant Analysis,‖ Analysis, 2010.

[30] P. Wu, H.-kai Peng, J. Zhu, and Y. Zhang, ―SensCare : Semi-Automatic Activity Summarization System for Elderly Care,‖ Scenario

Paper ID: NOV163314

890