Efficient TV White Space Database Construction via Spectrum Sensing and Spatial Inference

Mengyun Tang*, Ze Zheng*, Guoru Ding*†, Zhen Xue*
*College of Communications Engineering, PLA University of Science and Technology, Nanjing, China
†National Mobile Communications Lab, Southeast University, Nanjing, China
Email: mengyun_tang@yeah.net, 18551741025@163.com, dingguoru@gmail.com, xzalways@sina.com

***Abstract*—This paper presents an efficient method to construct white space database for devices to communicate in TV white space (TVWS). The goal is to build a TVWS database which senses the spectrum signal strength from white space devices (WSDs). Considering the incompleteness of measurement data, we formulate the problem of spatial inference as a matrix completion problem and propose a data recovery method by combining a fixed point continuation algorithm (FPCA) with a popular k-nearest neighbor (KNN) algorithm. Simulation results show that the proposed approach has a better performance in the TVWS database recovery than the traditional FPCA.**

***Index Terms*—white space, propagation models, FPCA, KNN**

I. Introduction

Along with the rapid development of telecommunications technologies, Next-generation Mobile networks are expected to achieve a 1,000-fold increase compared to the current wireless network in terms of capacity [1]. The main incentive is expected to come from the improved network architecture, and the convergence of information and communication. To offer faster network speed and more diverse categories of business, we also focus on acquiring new spectrum and new wireless access technology, among which, the spectrum utilization is increased by sharing more unlicensed spectrum. In fact, the spectrum resources for allocation are almost being exhausted because of our growing demand. The lack of wireless band poses a challenge, which needs to be solved urgently.

The spectrum bands become highly under-utilized since the evolution of TV broadcasting, from analog into digital, as well as the improvement of video compression technologies. TV white space (TVWS) is the TV spectrum bands which is unoccupied by TV signals at certain time [2], ranging from 470MHz to 790MHz. Furthermore, a large-scale measurement, carried out across 30+ diverse locations in a typical metropolis, reveals that more than 50% and 70% of the TV spectrum are white spaces in outdoor and indoor scenarios respectively [3]. Recently, it has been a research hotspot to study efficient approaches of using these released TVWS efficiently.

So far, a series of detection algorithms have been investigated to acquire white space information. There are two schemes widely accepted to help white space devices (WSDs) obtain knowledge of channel occupancy [4]. One is spectrum sensing which identifies the status of the frequency band by using spectrum sensing algorithms [5]. In this way, the current channel is regarded as occupied when the detected signal strength is higher than the detection threshold; otherwise, the channel is available for WSDs.

The other scheme is geo-location database, which combines terrain data and propagation models to determine the max transmission power of each channel. According to the Ofcom, when we acquire TVWS by geo-location database, there are two categories in the system, master users and slave users [6]. The master users need to be positioned and have the ability to communicate to the database directly. The slave users do not have the positioning capability, hence they must be under the control of master users. In order to manage the unused licensed white space, among others, there are eight well-known geo-location databases being in use: Spectrum Bridge, Fairpectrum, NICT, Nominet, Google, Sony, iconectiv, and Microsoft [7]. The geo-location database scheme can provide an effective and technically feasible service for unlicensed devices to inquire the TVWS availability.

However, there are some limitations if we rely on the geo-location database absolutely. Usually, propagation models are not accurate enough but computationally expensive. Besides, the prediction of the availability of TVWS is dependent on the resolution of the terrain data, and low location granularity may result in the loss of white space. Motivated by these observations, this paper aims to construct a TVWS database via spectrum sensing and spatial inference. Specifically, through combining the database and the sensing scheme, we propose an easy implemented approach to learn the TVWS's availability from massive WSDs, and the main contributions throughout this paper are summarized as follows:

- Combine the geo-location database and spectrum sensing scheme, in which we regard the signal strength calculated by propagation model as the ground truth and use the sensing data as the sampled matrix to be completed.
- Introduce gridding into the construction of the TVWS database. It makes data analysis and processing more feasible by dividing the area into a large number of small grids.
- Propose an improved FPCA algorithm with KNN. KNN is exploited to complete the unknown grid when the number of the neighbors is enough to be recovered.
- Provide simulations to present the effectiveness of the improved FP-KN over the traditional FPCA, the proposed algorithm has a better completion performance.

The remainder of the paper is organized as follows. Section II presents the system model and the process of obtaining TVWS information. Section III summarizes different kinds

978-1-4673-8590-9/15/$31.00 ©2015 IEEE

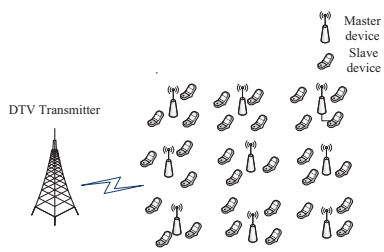


Fig. 1. TV transmitter and its coverage network where massive devices are inside

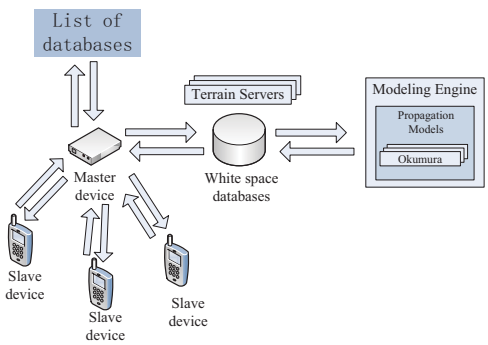


Fig. 2. Devices acquire TVWS by geo-location database

of propagation models and emphasizes the model used in this paper. An improved solution to fixed point continuation algorithm is provided in Section IV. Section V presents the performance evaluation and tests the proposed algorithm using the UHD (USRP hardware driver) platform. The conclusions are drawn in Section VI.

II. System model

In this section, we consider a scenario where the DTV broadcasting network shares channels with the WSDs [8]. Generally, high power TV transmitter can cover a vast area which extends to hundreds of kilometers. Users can be closely located anywhere in its coverage area, and each of them has a receiver which can indicate the signal strength. It is illustrated in Fig. 1 that the network is divided into dozens of grids, where there is a master device located inside with massive slave devices around it.

The process of the geo-location database obtaining TVWS information is illustrated in Fig. 2. The master users send a request to the Ofcom and wait for the response when they need a DTV channel for communication [9]. Then they select a database from the Ofcom list of geo-location databases and send their own parameters. The database reply the information such as which channel they can use and what the limited power is through the complex calculation of terrain data and propagation models. At the same time, the slave user confirm these details by communicating to the master user.

III. Signal propagation models

Appropriate statistical wave modeling is required when the white space spectrum information is provided by the database. The main job of propagation models is to estimate the strength of received signal when given a transmitter's location, an antenna's height, transmit power and so on.

To perform a propagation model, we need to figure out how many areas WSDs can cover, and how much effective terrain data can be provided in practice. Most of the terrain where WSDs are located in may be irregular because there are mountains and valleys in the coverage area, as well as other man-made obstacles. All of these uncertainties would make the propagation modeling much more challenging [10].

Nowadays, there are various propagation models for UHF signals with varying complexity and accuracy [11]. Models, like Free Space, Egli, are simple to estimate signal strength with a few parameters. Here, we talk more about the following three more complex models which are commonly used in practical use.

- Longley-Rice [12] (L-R) model. L-R model uses route geometry of landform to calculate the transmission loss. necessary parameters in the formula include frequency, irregular terrain data, surface refraction and so on. It also introduces two parameters to indicate the media characteristics: permittivity and soil conductivity. In L-R model, there are totally 7 climate permittivities taken into account.
- F-curves. According to the FCC report, F-curves, a particular propagation model, is used by several approved spectrum databases [13]. It defines the protection contours of the TV channel. The establishment of F-curves model is based on specific measured data, such as operating band, effective radiated power (ERP), antenna beam parameters and so on. F-curves is required to estimate the contours of each radiated direction and the signal strength at the boundary curves should be 41 dBu. By using the average terrain to replace the dense terrain information, the contours can be estimated accurately.
- Okumura measurement. In [14], WSDs can be divided into two categories, one is fixed WS devices which are highly positioned and permitted high EIRP, while the other is portable WS devices with low power which are lowly positioned. NICT therefore proposed the following two corresponding path-loss models according to this classification. The signal attenuation model of the portable WS devices is simple because the locations are mainly near the ground and their distance is usually at short range. However, the fixed WS devices are needed to apply a more complex curve-based model called Okumura measurements.

In this paper, we take Okumura measurement provided by NICT as the simulation propagation model.

$$L_p = L_{free} + a_{mu} - h_{tu} - h_{ru}, \tag{1}$$

$$L_{free} = 20\log_{10}d + 20\log_{10}f_c + 32.45, \tag{2}$$

where we set TV channel center frequency f_c 615 MHz, and d is the Tx-Rx distance in kilometers. L_{free} and L_p represent the free-space attenuation and the urban path-loss, respectively. Additionally, a_{mu} is a set of curves representing the median attenuation related to L_{free}, and the set of correction curves h_{tu} and h_{ru} are provided by the Tx antenna effective height h_t and the Rx antenna effective height h_r.

We take the measurements estimated by propagation model as the ground truth, while master devices acquire the strength can be given as a Gaussian distribution, i.e.

$$\hat{P}_{i,j} \sim N\left(P_{i,j}, \frac{(P_{i,j}+N_0)^2}{N_{sam}}\right), \tag{3}$$

where $P_{i,j}$ indicates the original signal strength, N_0 is the noise power, and N_{sam} is the number of known grids.

The signal strength of a lot of unknown locations cannot be acquired if there are not adequate WSDs, but excessive WSDs will cause the waste of the resource and increase the amount computation. In this paper, We move the massive data processing to the database engine from the WSDs which are not good at processing. We propose an effective solution to recover the signal strength with few WSDs because of the sparse and discontinuity characters of the collected strength matrix.

IV. Recovery improved algorithm

In the spectrum sensing based technique, we can apply an effective solution which is called fixed point continuation algorithm (FPCA) [15]. First, we divide the explored white space area into a set of small square grids. We assume that the dataset in each grid is a $p \times m$ matrix.

A portion of the spectrum measurement results of the grids can be collected according to a given sampling rate, denoted with the sensed average PU signal strength. For example, the spectrum measurement result of the grid which locates in (i,j) can be indicated as $M_{i,j}$. The other grids would be set as "unknown". After a period of sensing observation, the subset matrix E can be defined as:

$$M^E_{i.j} = \begin{cases} M_{i,j}, & (i,j) \in E \\ 0, & \text{otherwise} \end{cases}. \tag{4}$$

The matrix completion for spectrum measurements of "unknown" grids without spectrum sensing can be modeled as completing the matrix using the known subset M^E , specifically implemented as the following optimal problem [16]:

$$\min_{M \subset R^{m\times n}} \nu||M||_* + \frac{1}{2}\sum_{(i,j)\in E}|M_{i,j} - M^E_{i.j}|^2, \tag{5}$$

where $||M^*||_*$ denotes the nuclear norm of matrix M (sum of singular values of M^* , ν is a scaling parameter which can balance the two arithmetic operations of summing.

Then we apply FPCA to solve this optimal problem. The key of FPCA algorithm is the arithmetic separation technique, which is based on the following fixed-point iterative linear calculation:

$$\begin{cases} Y^k = M^k - \tau g\left(M^k\right) \\ M^{k+1} = S_\nu\left(Y^k\right) \end{cases}, \tag{6}$$

in which $S_\nu(\cdot)$ is the matrix shrinkage operator. τ is the iterative operator. $g\left(M^k\right) = A^*\left(A\left(M^k\right) - M^E\right)$ is the gradient of function $\frac{1}{2}||M_{i,j} - M^E_{i,j}||^2_2$ at the point M^k and A is a linear operator $A(M^k) = A_1M_1 + A_2M_2 + \ldots + A_kM_k$ indicating a linear mapping from R^k to $R^{m\times n}$. We notice that the above equations are actually the optimality condition for the following convex problem:

$$\min_{x^*} \nu||M^*||_1 + \frac{1}{2}||M^* - M^E||^2_2. \tag{7}$$

This convex problem has a closed optimal solution given by the shrinkage operator $S_\nu(\cdot)$ [16]:

$$M^* = \widetilde{S}_\nu(M^E). \tag{8}$$

When the shrinkage operator $S_\nu(\cdot)$ is determined by $\mathrm{sgn}(\cdot) \odot \max\{|\cdot| - \nu, 0\}$, the fixed point iterative algorithm can be given by:

$$M^{k+1} = \widetilde{s}_\nu(M^k - \tau g^k). \tag{9}$$

Using this recursion algorithm, the optimum solution can be finally obtained as $M* = S_\nu(I(\cdot) - \tau g(\cdot))$ and it is exactly the matrix reconstructed from its sampling known elements.

However, FPCA, which is usually used to tackle low-rank matrix recovery problem, may not be appropriate for the clustered signal through the noisy DTV channel. The signal strengths of data matrix do not have the nature of low rank property. Here we propose an approach which combines FPCA with another algorithm KNN to improve the performance of recovery, which we defined it FP-KN.

K-nearest neighbor (KNN) is one of the most popular machine learning algorithms [17]. In this paper, we take KNN as a matrix recovery algorithm which completes the unknown grid with the neighbor data around it. The main idea of KNN reconstruction is to calculate the mean of the neighbors of different distances, and the performance of recovering depends on the numbers of neighbors. KNN is relatively accurate because of the characteristic of utilizing the nearest neighbors. However, when the matrix is so sparse that some unknown grids have no neighbors, KNN becomes useless.

The FP-KN algorithm implements KNN to recover the matrix for the first step when the unknown grids have enough neighbors. It will search for the grids that has more than one neighbors and calculate the weighted mean of the neighbors as an alternative of the current measurement. Then FPCA algorithm will be used to continue completing the reconstructed matrix. By using this twofold recovery method, we can achieve a better completion performance.

V. Performance evaluation

A. Simulation Setup.

We consider a DTV broadcast environment that WSDs can reuse a TV channel licensed to a large-scale DTV system. In the following simulations, we set the center frequency of TV

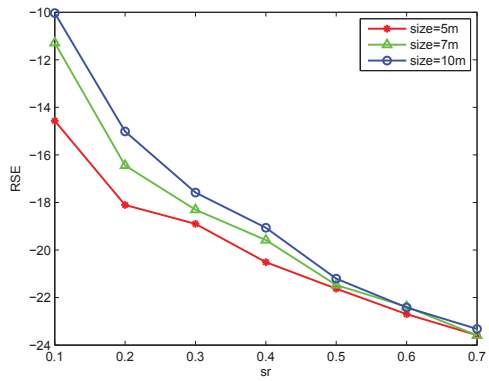


Fig. 4. RSE performance of FPCA under different grid sizes

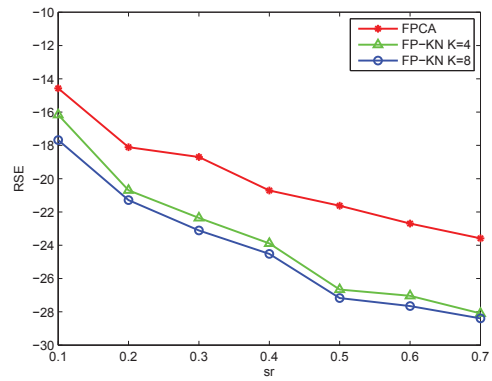


Fig. 5. RSE performance of FPCA and KNN under different sampling rate levels

channel f_c is 615 MHz and the transmission power of a DTV transmitter P_t is 10^6 $watt$. The parameters of propagation model are mainly based on the specifications in the Okumura measurement model.

In this section, the experiments are performed to evaluate the matrix completion algorithm we proposed above. The area of interest is divided into 100×100 small grids where there are three TV transmitters located inside, and each grid is $5km$ $\times$ $5km$ wide. Suppose that, the observed matrix is randomly sampled with the sampling rate sr in our experiments, where sr indicates the the percentage of known elements in the total spectrum data matrix M.

B. Simulation Results

To make our work more familiar, Fig. 8 presents the results of the completion formulation of FPCA and FP-KN. In this simulation , we first generate the ground-truth DTV coverage (see Fig. 8(a)) calculated by propagation model. Then, in Fig. 8(b), considering only 50% of of the locations being sampled by WSDs, the spectrum measurements are collected by master devices. The DTV coverage is quite difficult to be recognized from the sparse and noisy spectrum data. The matrix completion algorithm FPCA is used to recover the unknown measurements (Fig. 3(c)). The reconstruction performance of the improved algorithm FP-KN is shown in Fig. 3(d).

In order to measure the recovery performance more rationally, we define root square error (RSE) as the evaluation norm.

$$RSE\,[dB] = 10\log_{10}\frac{||\tilde{M} - G||_2}{||G||_2}, \tag{10}$$

in which $\tilde{M}$ is the reconstructed matrix and G is the ground truth calculated by the Okumura propagation model.

The recovery error generally decreases with an increasing sampling rate, as well as the decrease of the grid size (the length of each grid). As Fig. 4 presents, when sampling rate is 0.5, the RSE has 4 dB better than the circumstance when sampling rate is 0.3. Hence, we can find that smaller grid size or higher spatial resolution yields better recovery performance.

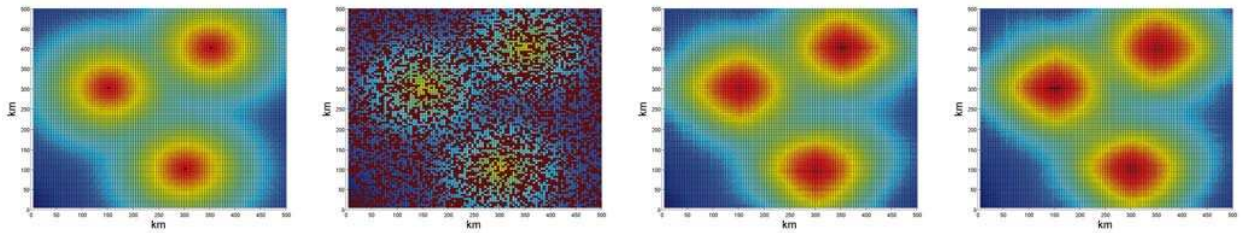
Fig. 5 shows that the root square error of spectrum sensing against different sampling rate for the geo-location database. When sampling rate is 0.3, the RSE of FP-KN (k=4) has 4 dB better than the ordinary FPCA algorithms. It is obvious that FP-KN has a better performance than FPCA. In addition, the RSE is smaller when the number k of the unknown grid's neighbors is larger, which means the reconstruction of the unknown grids is more accurate when k is lager with KNN algorithm.

C. Real-World Measurements and Analysis

To test the proposed algorithm in an indoor environment, we deploy an experiment consisting of a transmitter, a USRP (Universal Software Radio Peripheral), and a laptop computer. The central frequency of the transmitter is 992.98MHz. Before the test, as fig. 6 shows, we divide the laboratory room into 23×14 grids. We measure the signal strength of each grid in the laboratory room using the USRP hardware. After the real-world spectrum map is plotted, we randomly sample the spectrum data with the sampling rate of 0.4, and reconstruct the matrix using the FP-KN proposed in this paper. In fig. 7 we can see the sparse spatial observations via random sampling where the dark grids are represented as the unknown grids. Fig. 8(a) shows the real-world measurements achieved in this experiment, while fig. 8(b) shows the completed spectrum data with the FP-KN algorithm. By the comparing the two results, we can see the performance of FP-KN algorithm for real-world measurement.

VI. Conclusion

In this paper, we combine the geo-location database and spectrum sensing method to construct a TVWS database. To achieve this goal, we formulate the problem of unknown measurements reconstruction as an optimization problem and propose an improved algorithm. This algorithm takes the advantage of both k-nearest neighbor algorithm and the fixed point continuation algorithm, which shows its well performance in incomplete sensing data recovery.



(a) Ground truth (b) Sparse spatial observations (c) Spatial inference via FPCA (d) Spatial inference via FP-KN

Fig. 3. The completion formulation of FPCA(sr=0.5)

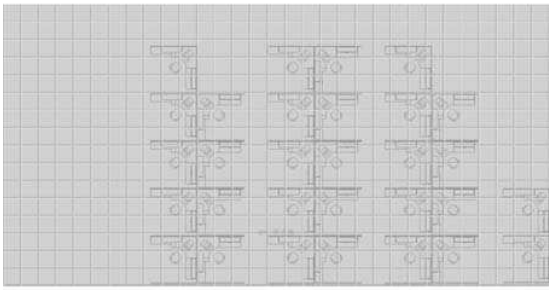


Fig. 6. Gridding of the laboratory room

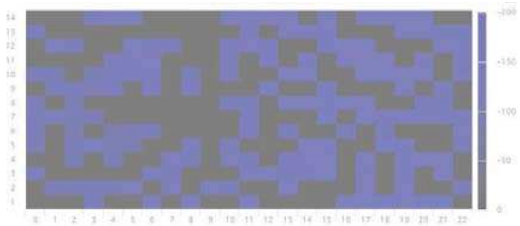
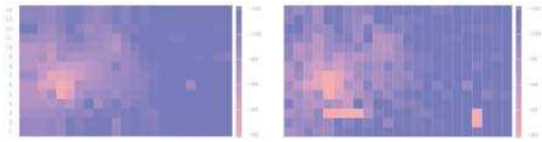


Fig. 7. Sparse spatial observations via random sampling



(a) Real-world measurements (b) Spatial inference via FP-KN

Fig. 8. The comparison of the reconstruction results

Acknowledgment

This work is supported by the National Natural Science Foundation of China (Grant No. 61501510 and No. 61301160), and Natural Science Foundation of Jiangsu Province (Grant No. BK20150717), and Jiangsu Planned Projects for Postdoctoral Research Funds.

References

[1] Q. Li, H. Niu, A. Papathanassiou, and G. Wu, "5G network capacity: key elements and technologies," *IEEE, Vehicular Technology Magazine*, vol. 9, no. 1, pp. 3129-3143, 2014.

[2] C. S. Sum, G. P. Villardi, M. A. Rahman, T. Baykas, H. N. Tran, Z. Lan, and H. Harada, "Cognitive communication in TV white spaces: An overview of regulations, standards, and technology," *IEEE, Communications Magazine*, vol. 51, no. 7, pp. 138-145. 2013.

[3] X. Ying, J. Zhang, L. Yan, G. Zhang, M. Chen, and R. Chandra, "Exploring indoor white spaces in metropolises," *Proc. the 19th Annual International Conference On Mobile Computing And Networking ACM*, pp. 255-266, Sep. 2013.

[4] W. Nan, G. Yue, and E. Barry "Database-augmented spectrum sensing algorithm for cognitive radio,". *Proc*. ICC, 2015

[5] G. Ding, J. Wang, Q. Wu, et al, "Robust spectrum sensing with crowd sensors," *IEEE Transactions on Communications*, vol. 62, no. 9, pp. 3129-3143, Sep. 2014.

[6] H. Oliver , "TV white spaces in the UK, and a major trial within the Ofcom TV white spaces pilot," *WInnComm-Europe 2014*. Rome, Italy, Nov. 2014.

[7] Ofcom. Qualifying White Space Databases. *https://tvws-databases.ofcom.org.uk/*

[8] G. Ding, J. Wang, Q. Wu, Y. D. Yao, F. Song, and T. A. Tsiftsis, "Cellular-Base-Station Assisted Device-to-Device Communications in TV White Space,"*IEEE Journal on Selected Areas in Communications*, Accepted to appear.

[9] Ofcom."Implementing TV White Spaces," Feb. 2015.

[10] J. J. Egli, "Radio propagation above 40 MC over irregular terrain," *Proc. the Institute of Radio Engineers*, vol. 45, no. 10, pp. 383-1391, 1957.

[11] R. Murty, R. Chandra, T. Moscibroda, and P. Bahl, "Senseless: A database-driven white spaces network," *IEEE Transactions on Mobile Computing*, vol. 11, no. 2, pp. 189-203, 2012

[12] W. Xu, X. Zhou, and L. Wu, "Equivalent scattering modeling of longley-rice," *Naval Aeronautical Engineering Institute*, vol. 41, no. 4, pp.42-44, 2011.

[13] A. Chakraborty, and S. R. Das, "Measurement-augmented spectrum databases for white space spectrum," *Proc. the 10th ACM International on Conference on emerging Networking Experiments and Technologies ACM*, pp. 67-74, Dec. 2014.

[14] G. P. Villardi, , H. Harada, F. Kojima, and H. Yano, "Multi-Level Protection to Broadcaster Contourand its Impact on TV White Space Availability," *IEEE Transactions on Vehicular Technology*. Apr. 2015.

[15] E. T. Hale, W. Yin, and Y. Zhang, "Fixed-point continuation for 1-minimization: Methodology and convergence," *SIAM Journal on Optimization*, vol. 19, no. 3, pp. 1107-1130, 2008.

[16] S. Ma, D. Goldfarb, L. Chen, "Fixed point and Bregman iterative methods for matrix rank minimization," *Mathematical Programming*, vol. 128, no. 1, pp. 321-353, 2011.

[17] H. Zhang, A. C. Berg, M. Maire, and J. Malik, "SVM-KNN: Discriminative nearest neighbor classification for visual category recognition," *IEEE Computer Society Conference*, Vol. 2, pp. 2126-2136, 2006.