IMAGE-BASED PM2.5 ESTIMATION AND ITS APPLICATION ON DEPTH ESTIMATION

Jian Ma, Kun Li, Yahong Han, Pufeng Du, Jingyu Yang*

Tianjin University, Tianjin 300350, China

ABSTRACT

Air pollution is still a big threat to human health particularly for developing countries. It is highly demanding to measure air quality with daily-used devices such as smartphones. On the other hand, it is difficult to estimate the scene depth under the foul weather using traditional vision-based methods. This paper proposes an image-based method for PM2.5 estimation by capturing a single image. We extract high-level features based on convolutional neural network (CNN) and learn the mapping between the features and PM2.5 by support vector regression (SVR). Given a captured image, we can estimate the PM2.5 value in real time. With the estimated PM2.5, we can estimate the depth of scene using sparse prior and non-local bilateral kernel. Experimental results demonstrate that the proposed method achieves the same accuracy of PM2.5 estimation as commodity measurement devices, and estimates the accurate depth information that is even better than the "ground-truth" captured by a laser in the no-haze condition.

***Index Terms*—** PM2.5, convolutional neural network (CNN), support vector regression (SVR), depth estimation

1. INTRODUCTION

With the development of industry, the air quality deteriorates day by day, especially in developing countries. Suspended particles in the air seriously affect our respiratory system, where PM2.5 (particle pollution matter with a diameter less than 2.5 micrometers) is an important measurement of air pollution indicators. PM2.5 refers to the air in the air aerodynamic equivalent diameter of less than or equal to 2.5 micron particles. It can suspend in the air for a long time, and the higher its concentration in the air content, the more serious air pollution is.

Existing methods [18, 16] measure the PM2.5 via special devices, such as Hanvon M1. However, it is not convenient for people to carry a special device everywhere and the price is a little high. It is highly demanding to measure air quality with daily-used devices such as smartphones.

Another difficulty under foul weather is to estimate the scene depth using vision-based methods. Estimating the depth of natural landscape from RGB images, even from a single monocular image, is a challenging task for computer vision. The image or video captured by a camera only partially records the real-world color projection on the 2D plane and does not directly reflect the 3D structure of the actual scene. Therefore, it is difficult for a computer to automatically and accurately comprehend the actual complicated scene from a photograph, thus hindering the application of the image and video information. Most traditional methods of depth estimation from a single image rely on markov random field (MRF) [14, 8]. With the development of deep learning, some methods use convolutional neural network to compute the depth information [9]. However, under foul weather, it becomes even more challenging since the image features in the captured image would be less distinguishable for vision-based depth inference.

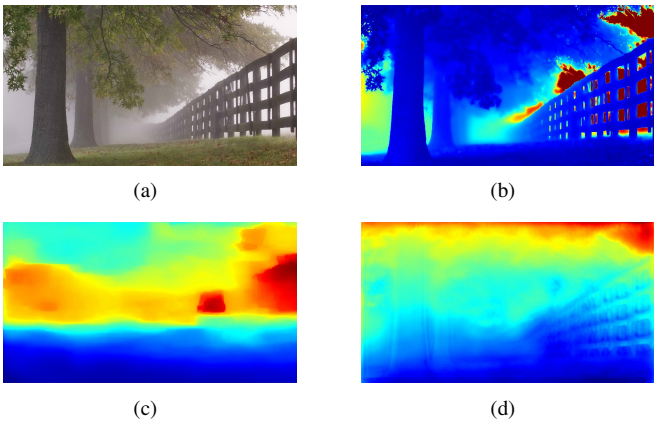


Fig. 1: PM2.5 estimation and depth estimation results: (a) a captured image under foul weather, (b) the recovered depth by our method with the estimated PM2.5 value ($209.507 \mu g/m^3$), (c) the recovered depth by method [7] and (d) the recovered depth by "hourglass" network [4].

In this paper, we propose a fast image-based method to estimate PM2.5 from a single captured image by daily-used devices such as smartphones. To our best knowledge, this is the first image-based method instead of any chemical or physical approach. With the estimated PM2.5, we also propose a depth estimation method with sparse prior and non-local bilateral kernel. Experimental results show that our method achieves fast and accurate estimation of PM2.5, and generates

*Corresponding author: lik@tju.edu.cn. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61571322 and Grant 61771339.

best depth estimation results compared with the state-of-the-art methods, even better than the "ground-truth" captured by a laser in the no-haze condition. Fig. 1 illustrates the depth estimation results under foul weather. Our code and dataset will be publicly available on the project website [1].

2. THE PROPOSED METHOD

We first propose a PM2.5 estimation method from a single image, which uses a convolutional neural network to extract haze-relevant features and SVR to learn the mapping between the feature and PM2.5. Further, using the estimated PM2.5, we recover the absolute depth through the atmospheric scattering model.

2.1. PM2.5 Estimation

2.1.1. Database

We capture 1575 images under various weather conditions using Iphone5s and measure the corresponding PM2.5 values using Hanvon M1. The images include normal natural images and sky images. Some captured images in our database together with the corresponding ground-truth PM2.5 values are given in our demo video [1].

2.1.2. Feature Extraction

We extract high-level features from an image using the VGG network [15] which is composed of convolution layers, pooling layers, full connection layers. Specifically, we resize the images into the size of 227×227 and then input them into the VGG network. After a series of layers, we extract a 4096-dimensional feature map from the fc7 layer.

2.1.3. Prediction

We learn the mapping f between the features and PM2.5 by the following support vector regression (SVR):

$$\begin{cases} \min\limits_{\omega,b} \frac{1}{2}||\omega||^2 + c\sum\limits_{i=1}^{m} \ell_{\in}(f(x_i) - y_i) \\ f(x_i) = \omega^T x_i + b \\ \ell(z)_{\in} = \begin{cases} 0, \text{if} |\text{z}| \leq \varepsilon \\ |z| - \varepsilon, \text{otherwise}, \end{cases} \end{cases} \tag{1}$$

where x_i is a feature of the i^{th} image, y_i is the ground-truth PM2.5 value of the i^{th} image, ω is the normal vector that determines the direction of the hyperplane, b is the displacement which determines the distance between the hyperplane and the origin, and c is a regularization constant. For m training samples, $\ell(z)_{\in}$ is an $\varepsilon-$insensitive loss function.

[1] http://cs.tju.edu.cn/faculty/likun/projects/PM2.5

2.2. Depth Estimation

The formation of an image is usually described as [10]:

$$I(x) = J(x)t(x) + A[1 - t(x)], \tag{2}$$

where I is the observed image which is photographed by users, J is true radiance of natural scene, A is the global atmospheric light composition, and t is the medium transmission. In general, the farther away from the camera, the thicker the haze, the more serious the impact of the object's reflected light by the suspended particles, the greater the proportion of global light to participate in imaging, resulting in the region with high brightness, texture blur and other features. Besides, the medium transmission map $t(x)$ describes the light portion that is not scattered and reaches the camera. By this prior, the depth of scene $d(x)$ can be estimated using the transmission map $t(x)$ and atmospheric scattering coefficient β:

$$d(x) = -\frac{1}{\beta} \ln t(x). \tag{3}$$

We propose a new transmission optimization method with sparse prior and non-local bilateral kernel by initializing using a dark channel method [5]. Specifically, we optimize the transmission map by minimizing the following function:

$$\sum_{x} \frac{\left(t(x) - \widetilde{t}(x)\right)^2}{\sigma^2(x)} + \lambda \sum_{x} \sum_{y \in N(x)} \sqrt{\alpha_{x,y}} \|t(x) - t(y)\|_1, \tag{4}$$

where λ is a penalization parameter balancing the importance between data term and smoothness term, $\sigma(x)$ is the standard deviation of $\widetilde{t}(x)$, $\|\cdot\|_1$ represents the ℓ_1 norm, $N(x)$ is the neighborhood of pixel x, and $\alpha_{x,y}$ is a pairwise weight between pixel x and pixel y, which is defined as follows:

$$\alpha_{x,y} = \exp\left(\frac{||\mathbf{B}_x{}^{\circ}(\mathcal{P}_x - \mathcal{P}_y)||_2^2}{\vartheta_1^2}\right), \tag{5}$$

where ϑ_1 controls the decay rate of the exponential function, $\mathcal{P}_x$ denotes an operator that extracts a $w \times w$ (3×3) patch centered at x, $\circ$ represents the element-wise multiplication. The bilateral filter kernel B_x is defined in the extracted $w \times w$ patch:

$$\mathbf{B}_x(x,y) = \exp\left(-\frac{||\overline{x} - \overline{y}||_2^2}{\vartheta_2^2}\right) \exp\left(-\frac{\sum\limits_{i \in c} ||I_x^i - I_y^i||^2}{\vartheta_3^2}\right), \tag{6}$$

where ϑ_2 and ϑ_3 are parameters of the bilateral kernel to adjust the importance of the spatial distance and intensity difference, respectively. In the equation, the distance of local patches and the similarity between the pixel x and every pixel y in the neighborhood of x are evaluated. ϑ_1, ϑ_2 and ϑ_3 are 3.0483, 1000 and 0.2 respectively in our experiment part.

Image Number	1	2	3	4	5	6	7	8	9	10	11	MAD
Linear	20.58	8.46	108.47	182.21	181.41	198.02	198.39	210.76	228.52	252.25	273.85	5.88
Polynomial	221.82	220.90	221.48	220.54	220.64	221.62	220.52	220.99	221.61	222.62	223.36	69.70
RBF	195.88	186.21	190.37	213.29	214.38	216.70	216.66	218.52	219.08	223.55	226.17	59.00
Ground Truth	6.40	16.75	102.76	178.06	185.81	191.19	201.13	217.90	228.66	256.46	280.70	

Table 1: PM2.5 estimation result ($\mu g/m^3$) with different kernel functions.

We use ℓ_1 norm to model the piecewise smoothness of the transmission map and use non-local prior weighted by a bilateral kernel on a larger neighborhood to fully exploit structural correlation. Through this method, the optimized transmission map is accurate without loss of smoothness.

In order to compute the β, we generate synthetic hazy images by adding artificial haze to the haze-free RGB images in the 3D video dataset [17], and then estimate the PM2.5 value for each synthetic hazy image by the proposed method. First of all, we manually set β and A, so we get the medium transmission $t(x)$. Meanwhile, clear images can be easily synthesized haze through the Equ.2. At this time, β which we have set is the ground truth. Therefore, we learn the relationship between PM2.5 and β by synthetic experiment. Through a lot of statistical experiments, we find that the relationship between PM2.5 and β is linear as shown in Fig. 2. When PM2.5 is higher, β becomes larger so that the image is blurred. The relationship between PM2.5 and β is fitted by least squares method:

$$\beta = |a\mathrm{PM2.5} + b|, \tag{7}$$

where a and b are the parameters. The learnt values of a and b are 0.02864 and -3.411, respectively.

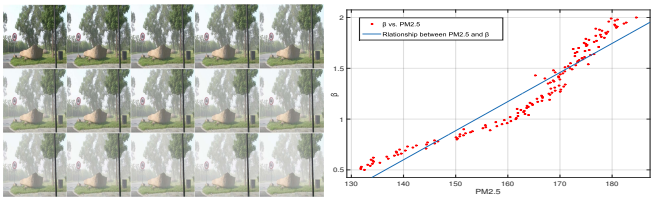


Fig. 2: Relationship between PM2.5 and β.

By this way, the depth map of scene can be estimated using Eq. (3) with the estimated $t(x)$ and β.

3. EXPERIMENTAL RESULTS

3.1. PM2.5 Estimation Results

We choose 1545 hazy images with different PM2.5 values as a training set, and 30 images as a validation set. We extract high-level features from an image using VGG-19 [15] via Caffe[6] which provides multimedia scientists and practitioners with a clean and modifiable framework for state-of-the-art deep learning algorithms and a collection of reference models. Then, the feature maps are trained by SVR by LIBSVM library [2]. Three kernel functions are tested: linear kernel function, polynomial kernel function, and Radial basis function (RBF). Table. 1 shows the PM2.5 results for randomly selected 11 images from the validation set, compared with ground truth measured by Hanvon M1. The mean absolute differences (MAD) is also reported in the table. Since the estimation accuracy could be significantly affected by the selection of kernel functions in SVR, we compare the results of three kernel functions, and the linear kernel function provides the far better estimation accuracy than the polynomial kernel function and the radial basis function (RBF). This suggests that the VGG network generates high quality features which are distinguishable in the feature space. Overall, our method provides quite promising prediction results particularly for high PM2.5 conditions.

3.2. Depth Estimation Results

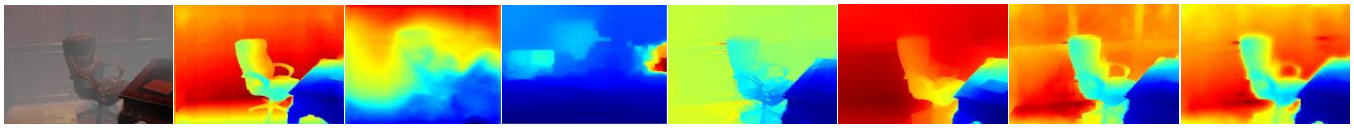
In this section, we evaluate the proposed depth estimation method on both synthetic dataset (Section 3.2.1) and real dataset (Section 3.2.2). The synthetic dataset is generated by adding different haze to the NYU-Depth v2 dataset [11] and the real dataset includes the public Make3D dataset [12, 13] and the images captured by users. We use following commonly-used measurements for quantitative evaluation:

- Relative error (Rel): $\frac{1}{T}\sum_p \frac{|d_p^{gt} - d_p^{est}|}{d_p^{gt}}$;
- Root mean squared error (RMSE): $\sqrt{\frac{1}{T}\sum_p \left(d_p^{gt} - d_p^{est}\right)^2}$;
- $\log_{10}$ error (log10): $\frac{1}{T}\sum_p |\log_{10} d_p^{gt} - \log_{10} d_p^{est}|$;

where d_p^{gt} and d_p^{est} are the ground-truth depth and the estimated depth at pixel p, respectively. T is the total number of pixels of the image.

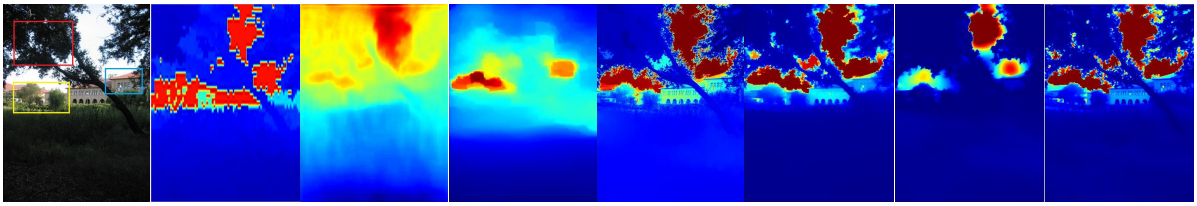
3.2.1. Results on Synthetic Dataset

We generate synthetic dataset by artificially adding haze on the haze-free images in NYU-Depth v2 dataset [11]. The NYU-Depth v2 dataset contains video sequences of various indoor scenes recorded by both the RGB and depth cameras of Microsoft Kinect. The synthetic hazy images are generated using the haze-free color images and its corresponding depth maps according to Eq.(2).



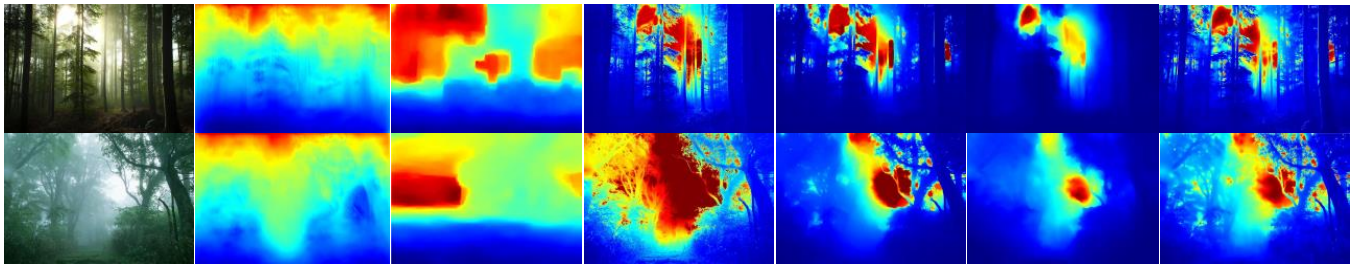
(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 3: Depth estimation results on (a) NYU synthetic dataset by methods (c) Chen *et al.* [4], (d) Karsch *et al.* [7], (e) He *et al.* [5], (f) Berman *et al.* [1], (g) Chen *et al.* [3] and (h) ours, compared with (b) ground truth. The PM2.5 value estimated by our method is 208.402 $\mu g/m^3$.



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 4: Depth estimation results on (a) Make3D dataset by methods (c) Chen *et al.* [4], (d) Karsch *et al.* [7], (e) He *et al.* [5], (f) Berman *et al.* [1], (g) Chen *et al.* [3] and (h) ours, compared with (b) ground truth. The PM2.5 value estimated by our method is 83.443 $\mu g/m^3$.



(a) (b) (c) (d) (e) (f) (g)

Fig. 5: Depth estimation results on (a) real images by methods (b) Chen *et al.* [4], (c) Karsch *et al.* [7], (d) He *et al.* [5], (e) Berman *et al.* [1], (f) Chen *et al.* [3] and (g) ours. From top to bottom, the PM2.5 values estimated by our method are 203.609 $\mu g/m^3$ and 284.328 $\mu g/m^3$, respectively.

In Table 2, we compare our result with five state-of-the-art methods. It can be seen that our method obviously outperforms other methods with smallest errors, which benefits from the elegant design of the optimization and the computation of β to get the absolute depth. Qualitative evaluation is given in Fig. 3. Our algorithm is fast: it takes 11s for 561×427 image size (including transmission estimation procedure).

Method	Rel	log10	RMS
Chen *et al.* [4]	3.6136	0.2135	0.2656
Karsch *et al.* [7]	1.5458	0.7160	0.5567
He *et al.* [5]	0.9163	0.1918	0.245
Berman *et al.* [1]	1.4717	0.1197	0.1515
Chen *et al.* [3]	0.5426	0.0978	0.1341
Our Method	**0.3308**	**0.0974**	**0.1129**

Table 2: Quantitative comparison with state-of-the-art methods. Lower is better.

3.2.2. Results on Real Datasets

We first evaluate the proposed method on public Make3D dataset [12, 13], compared with five state-of-the-art methods in Fig.4. It can be seen that our method generates the best depth map even better than the ground truth captured by a 3D laser, especially for the regions marked by color rectangles. Our algorithm takes 7s for 343×458 image size.

We also compare the depth estimation results for real images downloaded from Internet or captured by a camera in Fig.5. Our method also shows the best performance for wild images. Please find more results in our demo video [1].

4. CONCLUSIONS

In this paper, we propose an image-based PM2.5 estimation method via deep learning, and also propose a new depth estimation method using sparse prior and non-local bilateral kernel. Experimental results show that our method achieves the same accuracy of PM2.5 estimation as commodity measuring device, and the estimated depth map is accurate, even better than the "ground-truth" captured by a laser.

5. REFERENCES

[1] D. Berman, T. treibitz, and S. Avidan. Non-local image dehazing. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[2] C. C. Chang and C. J. Lin. LIBSVM: A library for support vector machines. *Acm Transactions on Intelligent Systems and Technology*, 2(3, article 27):389–396, 2007.

[3] C. Chen, M. N. Do, and J. Wang. Robust image and video dehazing with visual artifact suppression via gradient residual minimization. In *Proc. European Conference on Computer Vision (ECCV)*, 2016.

[4] W. Chen, Z. Fu, D. Yang, and J. Deng. Single-image depth perception in the wild. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 730–738. Curran Associates, Inc., 2016.

[5] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2341–2353, 2010.

[6] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. *Eprint Arxiv*, pages 675–678, 2014.

[7] K. Karsch, C. Liu, and S. B. Kang. Depth extraction from video using non-parametric sampling. In *Proc. European Conference on Computer Vision (ECCV)*, pages 775–788, 2012.

[8] B. Liu, S. Gould, and D. Koller. Single image depth estimation from predicted semantic labels. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1253–1260, 2010.

[9] F. Liu, C. Shen, G. Lin, and I. Reid. Learning depth from single monocular images using deep convolutional neural fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(10):1–1, 2015.

[10] W. E. K. Middleton. Vision through the atmosphere. *Handbuch Der Physik*, 1952.

[11] P. K. Nathan Silberman, Derek Hoiem and R. Fergus. Indoor segmentation and support inference from RGBD images. In *Proc. European Conference on Computer Vision (ECCV)*, 2012.

[12] A. Saxena, S. H. Chung, and A. Y. Ng. Learning depth from single monocular images. *Advances in Neural Information Processing Systems*, 18:1161–1168, 2005.

[13] A. Saxena, S. H. Chung, and A. Y. Ng. 3-D depth reconstruction from a single still image. *International Journal of Computer Vision*, 76(1):53–69, 2008.

[14] A. Saxena, M. Sun, and A. Y. Ng. Make3D: Learning 3D scene structure from a single still image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(5):824–840, 2008.

[15] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *Computer Science*, 2014.

[16] Z. Tao, D. Liu, Z. Wang, X. Ma, Q. Zhang, C. Xie, G. Bo, S. Hu, and Y. Wang. Measurements of aerosol phase function and vertical backscattering coefficient using a charge-coupled device side-scatter lidar. *Optics Express*, 22(1):1127–34, 2014.

[17] G. Zhang, J. Jia, T. T. Wong, and H. Bao. Consistent depth maps recovery from a video sequence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(6):974–988, 2009.

[18] W. Zhang, S. Gao, X. Song, J. Liu, W. Liu, and Z. Chen. Concentration measurement and metrical technology of fine particulate matter pm_(2.5). *China Powder Science & Technology*, 2013.