Pacific Graphics (2014)
J. Keyser, Y. J. Kim, and P. Wonka (Editors)

Example-based Haze Removal with two-layer Gaussian Process Regressions

Xin Fan, Renjie Gao and Yi Wang

School of Software, Dalian University of Technology

Abstract

Hazy images suffer from low visibility and contrast. Researchers have devoted great efforts to haze removal with the prior assumptions on observations in the past decade. However, these priors from observations can provide limited information for the restoration of high quality, and the assumptions are not always true for generic images in practice. On the other hand, visual data are increasing as the popularity of imaging devices. In this paper, we present a learning framework for haze removal based on two-layer Gaussian Process Regressions (GPR). By using training examples, the two-layer GPRs establish direct relationships from the input image to the depth-dependent transmission, and meanwhile learn local image priors to further improve the estimation. We also provide a method to collect training pairs for images of natural scenes. Both qualitative and quantitative comparisons on simulated and real-world hazy images demonstrate the effectiveness of the approach, especially when white or bright objects and heavy haze regions appear and existing dehazing methods may fail.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation— I.3.m [Computer Graphics]: Computational Photography—Image Processing

1. Introduction

Images of outdoor scenes are typically degraded by haze, showing low visibility and contrast. These degraded images bring great difficulties to either human perception or automatic computer vision systems. Researches have invested significant efforts to haze removal (dehazing) in order to restore images with high quality and visibility [Fat08,HST12]. We address this issue from a learning perspective that trains regressors on both imaging models and image priors from examples for image dehazing.

In the past decade, the dehazing algorithms from a single image have attracted great attention. Typically, these methods employ strong image priors for the recovery. Tan [Tan08] assumes that haze scenes are of high smoothness and consistence except for depth boundaries, and restores color contrasts by maximizing the local contrast difference. Fattal [Fat08] divides the scene into foreground objects and background shading under the assumption that the transmission is independent of surface shading. Yan [YXJ13] and Matlin [MM12] denoise the image prior for haze removal and underwater scenes. He *et al.* [HST12] propose the dark channel prior (DCP) that the minimum of RGB channels for a local patch is close to zero without haze. Therefore, the local minimum can be used to estimate the transmission. The calculation of DCP is simple yet effective, which gains great success for haze removal. There exist many variants of DCP to improve the performance from various aspects. Lv *et al.* [LCfS10] refine the transmission with bilateral filters and implement parallel processing on GPUs. Gao *et al.* [GFZL12] propose a transmission model derived from DCP that is able to edit the amount of haze in an image. However, the DCP assumption does not hold on sky and bright white regions that are quite common in outdoor scenes. Significant distortions inevitably appear in these scenarios.

These priors for either transmission or images are built upon the field knowledge of the designers on haze observations, and are not necessarily true for generic images. Researchers resort to image priors learnt from examples. In a pioneering work [FPC00], Freeman *et al.* learn local image priors from training examples with Markov random fields (MRFs). Recently, Gibson *et al.* [GBN13] combines the DCP based transmission model with the patch-wise MRFs for haze removal on synthetic images, showing promising results. Restoration algorithms are also able to take the advantage of learning image formation models from exam-

DOI: 10.2312/pgs.20141260

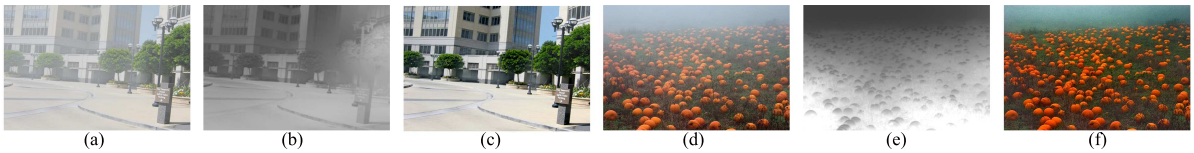

Figure 1: ***Haze removal of single image using our method:*** *(a)Input hazy image (simulated haze). (b) Estimated transmission map of (a). (c) Our haze removal result of (a). (d)Input hazy image (real haze). (e) Estimated transmission map of (d). (f) Our haze removal result of (d).*

ples in addition to the learnt priors for local patches. Tang *et al.* [TYW14] investigate haze-relevant features in a learning framework for haze removal, where the random forrest is applied to transmission learning. Their work relies on the DCP based features, neglecting the structural ones that also imply the depth-dependent transmission. The priors on local image patches important for high quality restoration are not included either.

In this paper, we propose an example based algorithm for haze removal with two-layer Gaussian process regressions (GPR). One layer of GPR is to build the regression between the transmission from rich features on super pixels [RM03] including color [TYW14], DCP [HST12], and structural [CLZ04] information. The other layer learns the priors on local structures for the predicted transmissions. The physical formation as well as local image priors are learnt from examples in a unified GPR framework. GPR implicitly encodes the smoothness as the MRF priors in [Tan08, FPC00, GBN13], but also explicitly bridges transmission and observations. These direct connects circumvent iterative inference like belief propagation for MRFs, and yield efficient estimation.

Moreover, it is a challenging issue to collect training examples of haze-free/hazy images and true transmission. We generate the training pairs on real-world images using the natural transmission model [GFZL12], differing from the 3D simulated haze [GBN13] and patch-wise simulated haze in [TYW14]. These natural examples is more likely to be applicable to images of outdoor scenes. Figure 1 demonstrate the estimated transmission and restored images on simulated and real-world scenic images.

2. Haze image formation model

We employ the formation model for a hazy image widely accepted in computer vision and computer graphics [Fat08, HST12]:

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

where I is the observed hazy image, J is haze-free scene, A is the atmospheric light, and t is medium transmission representing the portion of the light that reaches the camera with a scalar value ($0 \sim 1$). The relationship between transmission t and depth d can be expressed as:

$$t = e^{-\beta d(x)}, \tag{2}$$

where β is the scattering coefficient of the haze.

We take the modified transmission model [GFZL12] that preserves the aerial perspective for natural haze generation. The ratio of the input transmission t_i to the desired t_j as the transmission t to be estimated:

$$t(x) = t_i(x)/t_j(x), \tag{3}$$

The desired transmission t_j can be derived as:

$$t_j(x) = t_i(x)^{Dvis_i/Dvis_j}, \tag{4}$$

where $Dvis_i$ and $Dvis_j$ are the maximum visibility of the observation and desired image, respectively. These two parameters are tunable in order to decrease and increase the amount of haze in an image. We employ the model to generate the training examples for learning GPR.

3. Two-layer Gaussian Process Regression

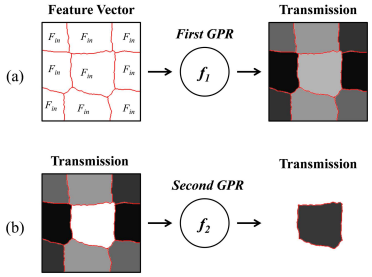

Figure 2: ***Learning framework with two-layer GPRs:*** *(a) the first layer GPR, inputs are multi-scale feature vectors and targets are transmissions; (b) the second layer GPR, inputs are the transmissions of neighbors and the target is the transmission at the super-pixel of interest.*

We give the learning based haze removal using two-layer GPRs on super-pixels. The first GPR layer takes multi-scale feature vectors as the input and transmissions as the target, and the second GPR uses the predicted neighboring transmissions as the input and the transmission at the super-pixel of interest as the target, as shown in Fig. 2.

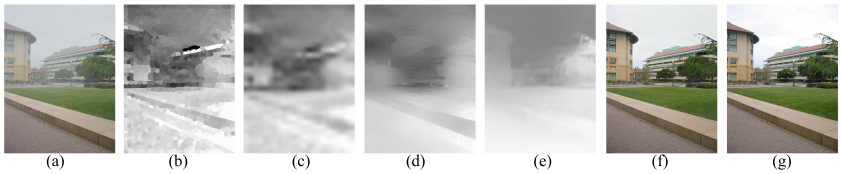

Figure 3: ***Haze removal with two-layer GPR:*** *(a) Hazy image. (b) Rough transmission estimated by the first layer GPR. (c) Refined transmission by the second layer GPR. (d) Final transmission improved by guided filtering [HST11] and Gaussian filtering. (e) Ground truth transmission. (f) Haze removal result. (g) Original haze-free image.*

3.1. Gaussian process regression form feature to transmission

We define the input of the first GPR layer $f_1(\cdot)$ as the average feature vector within the super-pixel of interest:

$$F_{in} = \frac{1}{|\gamma|} \sum_{p_i \in S_i} F(p_i), \tag{5}$$

$$F(p_i) = \left[H, D^4, C^4, S^4, G^{3,8}\right] \tag{6}$$

where S_i is the super-pixel, γ is the number of pixels in S_i, and $F(p_i)$ is the multi-scale feature extracted at the pixel p_i. The feature vector includes H, D^4, C^4, S^4 and $G^{3,8}$, representing the hue, DCP with 4 scales, contrast with 4 scales, saturation with 4 scales and gabor features with 3 scales and 8 orientations, respectively (details can be found in Section 4.2). The recent work [TYW14] construct a 325-dimensional feature vector at every pixel in a 5×5 patch. We take the average of the pixel feature vectors within a super-pixel where these vectors share similar characteristics. The average significantly decrease the regression complexity.

It is widely accepted that the estimation of the transmission map is essential to haze removal [HST12, GVN12, GFZL12, TYW14], and transmissions of an image are locally constant. We take the transmission at a super-pixel as the target shown in Fig. 3 as transmissions for the first layer of GPRs. It is nontrivial to collect the target transmission values as it is not directly available from an observation. We employ the natural transmission model in [GFZL12] to simulate haze with specific transmissions, which we detail in the next section, and then feed the pairs of transmissions and feature vectors extracted from the simulated images to train the parameters of the GP regressor.

Given a trained GPR, we simply take the mean of $p(y_p|y)$ as the predicted transmission at every super-pixel by substituting the feature vectors of an observation into. The predicted transmission is able to roughly reflect the depth and global structures of an image shown in Fig. 3(b), but present evident discrepancy on neighboring super-pixels. Therefore, it is necessary to apply one more GPR layer in order to impose the local smoothness to the predicted transmission map.

3.2. Gaussian process regression on neighboring transmissions

We use the second GPR layer to smooth the local discrepancies among predicted transmissions. These discrepancies destroy the consistency of image structures, and may lead to unnatural estimation in local regions. This GPR layer plays an similar role to MRFs in [FPC00, GBN13], but needs no iterative inference process. We can achieve the smoothing by a one-step prediction.

We concatenate the predicted transmissions $\tilde{t}$ of the nearest 8 neighbors $\mathcal{N}(S_i)$ of a super-pixel S_i as the input:

$$\tilde{\mathbf{t}}_i = [\tilde{t}(S_1), ..., \tilde{t}(S_j), ..., \tilde{t}(S_8)]_{S_j \in \mathcal{N}(S_i)}. \tag{7}$$

It should be noted that we do not use the adjacent neighbors as they may yield different input dimensions. The target is the transmission $t(S_i)$ at current super-pixel S_i. We train the second GP regressor by using the transmissions of simulated hazy images and the predicted transmissions of the first GPR layer for the training. In the prediction stage, we feed $\tilde{\mathbf{t}}_i$ to the trained GP regressor, and take the mean of the conditional distribution as the smoothed transformation map.

The output of the second GPR layer shows global smoothness and consistency as illustrated in Fig. 3(c). For a further refinement, the guided filtering [HST11] is applied to obtain the final transmission (Fig. 3(d)). Our final estimation has little difference with the ground truth in Fig. 3(e).

4. Training examples for Gaussian process regressions

We present the process to collect the training examples for the two-layer GPRs including the generation of transmissions, over-segmentation for super-pixels and the construction of the multi-scale feature vectors.

4.1. Training data

Our training images with known transmissions are generated by two methods: (1) simulating hazy images by known depth on haze-free images; and (2) estimating transmission maps from hazy images. We apply both methods to *real-world natural* images in contrast to 3D scenes in [GBN13] and patch-wise simulated images in [TYW14]. It is more likely for our model to apply to real world images.

The first method is motivated by the depth learning of depth. In [SCN08], Saxena *et al*. use a 3D scanner to collect images and corresponding depth maps. Following (2), we convert depth maps into transmissions under the assumption of β constant to depth. We set a small positive constant (0.1) as the transmission corresponding to the maximum depth, and hence normalize the transmission maps. Subsequently, we apply the guided filtering to refine the transmissions [HST11]. Finally, we obtain hazy images through (1). We apply 4 transmission-based dehazing methods [Fat08, Tan08, HST12, GFZL12] to hazy outdoor images, and pick out the transmission maps that produce the best restoration among the four algorithms. We collect these transmission maps along with the hazy images as part of our training examples.

4.2. Super-pixels

We employ the super-pixels in [RM03] to assemble feature vectors. The super-pixels are obtained by using the following features:

1.inter-region texture similarity;

2.intra-region texture similarity;

3.inter-region brightness similarity;

4.intra-region brightness similarity;

5.inter-region contour energy;

6.intra-region contour energy;

7.curvilinear continuity.

This combination of features segments super-pixels reasonably. The pixels in a super-pixel present homogenous texture, brightness and color. The use of these mid-level features other than image patches renders lower dimensional input space that favors smaller training examples and simpler regressors, and meanwhile it partially provides structural contexts.

5. Multi-scale feature vector

The ability to incorporate multiple features is one of the advantages of regression methods. We list the features for our GPR training including the DCP based ones [TYW14]. First, we introduce the hue, multi-scale dark channel, local maximum contrast and saturation.

Hue disparity can be used to detect haze [AAHB11]. Therefore, in our work, we apply the hue channel of the image in Lch color space directly. The multi-scale dark channel, local maximum contrast and saturation can be expressed as:

$$D^r(x) = \min_{y\in\Omega_s(x)}\Big(\min_{c\in\{r,g,b\}}(I^c(y))\Big), \tag{8}$$

$$C^s(x) = \sqrt{1-\frac{(\max_{y\in\Omega_s(x)}(\max_{c\in\{r,g,b\}}(I^c(y))))^2}{(\min_{y\in\Omega_s(x)}(\min_{c\in\{r,g,b\}}(I^c(y))))^2}}, \tag{9}$$

$$S^s(x) = 1-\frac{\max_{y\in\Omega_s(x)}(\max_{c\in\{r,g,b\}}(I^c(y)))}{\min_{y\in\Omega_s(x)}(\min_{c\in\{r,g,b\}}(I^c(y)))}. \tag{10}$$

where D_r is the dark channel, I^c are colors of the original image I, and Ω_s is a local window centered at x with size $s\times s$. In this paper, we use four scales as $D^4=[D_1,D_4,D_7,D_{10}]$, $C^4=[C_1,C_4,C_7,C_{10}]$ and $S^4=[S_1,S_4,S_7,S_{10}]$.

Then, we apply gabor feature [CLZ04] to extend the information content of feature vector. Gabor filter outperforms in the field of textural orientation, which can be apparently changed by haze. The gabor wavelet transform can be defined as:

$$G^{i,j} = \sum_{s1}\sum_{s2} I(i-s1,j-s2)\psi^*_{i,j}(s1,s2), \tag{11}$$

where i and j are the scale and direction of the wavelet. $s1$ and $s2$ define the filter mask size, and $\psi^*_{i,j}$ is the complex conjugate, generating from dilation and rotation. More details can be found in [CLZ04].

6. Experiment results

In this section we mainly make comparisons as following: (1) the effect of the second layer GPR, compare our approach with the implement lack of the second layer GPR; (2) the advantage of super-pixels, compared with the usage of patches; (3) the effect of different features, the comparisons among $G^{3,8}$, $HD^4C^4S^4$ and $HD^4C^4S^4G^{3,8}$; and (4) the comparisons with existing methods in [Fat08, HST12] and the work of Photoshop, which is a widely used commercial software of image processing. In the experiment, we use 40 images as training set.

6.1. Two-layer Gaussian Process Regression *vs.* Single layer Gaussian Process Regression

We make comparisons between the processes with the two-layer GPR and with the single first layer GPR, meanwhile, keep other setting all the same, to show the importance of the second layer GPR. Fig. 4 shows the difference applying the double or single layer GPR. In Fig. 4(f), the color distortion of red rectangle regions is caused by the uneven distribution of transmission in sky regions. Besides, the obtained transmission without the second layer refined can have bad consistency in local details, as the left haze (the yellow rectangle region) shown in Fig. 4(f), the green leaves are restored well but the shadow not. Though these differences can be hardly distinguished from the transmission maps(Fig. 4(d) and (e)), it affects the result a lot. The application of the second layer GPR can help to improve the smoothness and consistency of transmissions, leading to good performance (Fig. 4(g)).

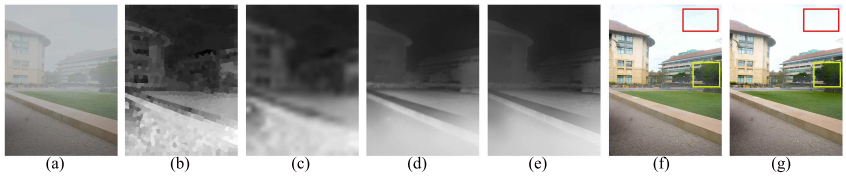

Figure 4: ***Comparisons between the two and single layer GPR:*** *(a) Hazy image. (b) Rough transmission estimated by the first layer GPR. (c) Refined transmission by the second layer GPR. (d) Final transmission generated by the single layer GPR. (e) Final transmission generated by the two-layer GPR. (f) Haze removal result generated by the single layer GPR. (g) Haze removal result generated by the double layer GPR.*

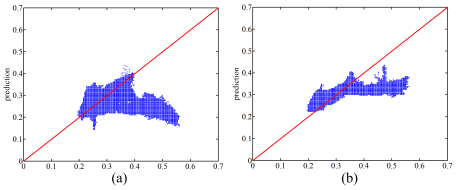

Figure 5: ***Predicted transmission vs. ground truth transmission on patches and super-pixels (for Fig. 6):*** *(a) Prediction of patches, MSE=1.41e-2. (b) Prediction of super-pixels, MSE=0.56e-2.*

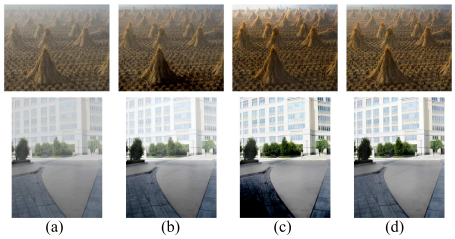

Figure 6: ***Comparison between different features:*** *(a)Hazy image. (b) Result of $G^{3,8}$. (c) Result of $HD^4C^4S^4$. (d) Result of $HD^4C^4S^4G^{3,8}$.*

6.2. Super-pixels *vs.* Patches

For comparisons between super-pixels and patches, we replace the super-pixels with 14×14 patches, which can generate similar numbers of sample for the testing image (400×533 with 1094 super-pixels). The difference mainly demonstrates at the general performance. Fig. 5 plots the distribution of prediction transmission vs. ground truth on patches and super-pixels, the prediction of patches is apparently lower than the truth and super-pixels' is much better. In addition to this, the comparison of MSE, 1.41e-2 of patches and 0.56e-2 of super-pixels, also proves the better performance of super-pixels.

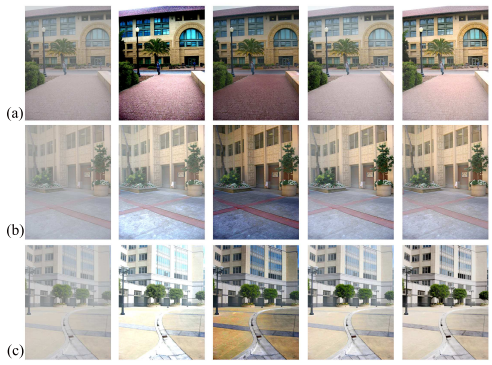

Figure 7: ***Comparisons with other dehazing methods (hazy images are estimated):*** *(a) Haze0_8, (b) Haze1_7, (c) Haze2_2. From left to right, hazy images, Fattal's results [Fat08], He* et al.*'s results [HST12], Photoshop's results and ours.*

6.3. Comparisons between different features

In the work of [TYW14], the feature of $D^4HC^4S^4$ proves to be the best among the combination of D, H, C, S respectively. Dark channel is the major feature and other features are assistant. This feature, however, contains only the local cues and lack of the texture information. As an improvement, we construct the similar $HD^4C^4S^4$ into the input feature vector and further add the $G^{3,8}$, the gabor feature with 3 scales and 8 orientations, with high weight. For proving the effectiveness of this combination, we mainly compare the features among $G^{3,8}$, $HD^4C^4S^4$ and $HD^4C^4S^4G^{3,8}$.

Fig. 6 describes the comparisons between different features. As a texture feature, the gabor features are not so closely associated with haze and the usage of single $G^{3,8}$ cannot restore well, both over-dehazing and under-dehazing may happen, as Fig. 6(b) shows. The $HD^4C^4S^4$ feature can dehaze well on images following the assumption of dark channel prior, like the image of cones in Fig. 6(c). Nevertheless, when the assumption is broken, like the white building with sunshine in Fig. 6(c), the prediction may fail, notice the over-dehazing dark ground. The combination of

$HD^4C^4S^4G^{3,8}$ can not only handle the images following dark channel prior well, but also restore the scene breaking the prior correctly due to the correction function of gabor features on dark channel, as shown in Fig. 6(d).

6.4. Comparisons with other dehazing methods

Finally, we compare our results with the existing methods in [Fat08, HST12] and the work of Photoshop. In the comparisons of data based on Saxena *et al.*'s work [SSN07], Fattal's [Fat08] work performs worst on visual effects (Fig. 7), either stretches contrast excessively or under-estimates the haze, due to the restrict assumption. In Fig. 7, The red-white ground (first row), yellow-white wall (second row) and sunny and white scene (third row) kill the assumption of dark channel, so serious color distortion occurs in He *et al.*'s results. Besides, the results are also too dim due to the usual over-dehazing of dark channel prior. The performance of Photoshop looks good, however, the results are obviously under-dehazed and big amount of haze is still remained. The original haze-free images are added with thin haze. Our work performs best both on visual effects, the estimated transmission is rather close to the real and the results are restored with high fidelity and visibility.

7. Conclusion and Discussion

In this paper, we demonstrate a two-layer GPR model to estimate the transmission and remove haze. By using training examples, the two-layer GPRs establish direct relationships from the input image to the depth-dependent transmission, and meanwhile learn local image priors to further improve the estimation. We also provide a method to collect training pairs for images of natural scenes. Experimental results presents that the advantage of our algorithm compared with other existing methods. Our model still has room for improvement and extension: (1) a faster algorithm for superpixels is expected to improve the efficiency of our algorithm; (3) the addition of other features, like the line space subdivision [BWW01] or view-dependent layered projective texture maps [RMD03], may improve the precision of regression further.

8. Acknowledgments

This work is partially supported by the Natural Science Foundation of China under grant Nos. 61033012, 11171052, 61272371, 61003177 and 61328206, the program for New Century Excellent Talents (NCET-11-0048). The author would like to thank the anonymous reviewers of this paper for many valuable comments that helped improve the quality of this paper. The author would also like to thank Jue Wang of Adobe research for discussing the algorithm development.

References

[AAHB11] ANCUTI C. O., ANCUTI C., HERMANS C., BEKAERT P.: A fast semi-inverse approach to detect and remove the haze from a single image. In *Proc. ACCV*. Springer, 2011, pp. 501–514. 4

[BWW01] BITTNER J., WONKA P., WIMMER M.: Visibility preprocessing for urban scenes using line space subdivision. In *Proc. PG* (2001), pp. 276–284. 6

[CLZ04] CHEN L., LU G., ZHANG D.: Effects of different gabor filter parameters on image retrieval by texture. In *Proc. ACM Int. Conf. Multimedia* (2004), Citeseer, pp. 273–278. 2, 4

[Fat08] FATTAL R.: Single image dehazing. *ACM Trans. Graph. 27*, 3 (2008), 72:1–72:9. 1, 2, 4, 5, 6

[FPC00] FREEMAN W. T., PASZTOR E. C., CARMICHAEL O. T.: Learning low-level vision. *Int. J. Comput. Vision 40*, 1 (2000), 25–47. 1, 2, 3

[GBN13] GIBSON K., BELONGIE S., NGUYEN T.: Example based depth from fog. In *Proc. of ICIP* (2013), IEEE, pp. 728–732. 1, 2, 3

[GFZL12] GAO R., FAN X., ZHANG J., LUO Z.: Haze filtering with aerial perspective. In *Proc. of ICIP* (2012), IEEE, pp. 989–992. 1, 2, 3, 4

[GVN12] GIBSON K., VO D., NGUYEN T. Q.: An investigation of dehazing effects on image and video coding. *IEEE Trans. Image Process. 21*, 2 (2012), 662–673. 3

[HST11] HE K., SUN J., TANG X.: Guided image filtering. *Proc. ECCV 35*, 6 (2011), 1397–1409. 3, 4

[HST12] HE K., SUN J., TANG X.: Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell. 33*, 12 (2012), 2341–2353. 1, 2, 3, 4, 5, 6

[LCfS10] LV X., CHEN W., FAN SHEN I.: Real-time dehazing for image and video. In *Proc. PG* (2010), pp. 62–69. 1

[MM12] MATLIN E., MILANFAR P.: Removal of haze and noise from a single image. In *IS&T/SPIE Electronic Imaging* (2012), International Society for Optics and Photonics, pp. 82960T–82960T. 1

[RM03] REN X., MALIK J.: Learning a classification model for segmentation. In *IEEE Int. Conf. Comput. Vis.* (2003), IEEE, pp. 10–17. 2, 4

[RMD03] RECHE-MARTINEZ A., DRETTAKIS G.: View-dependent layered projective texture maps. In *Proc. PG* (Oct 2003), pp. 492–496. 6

[SCN08] SAXENA A., CHUNG S. H., NG A. Y.: 3-d depth reconstruction from a single still image. *Int. J. Comput. Vis. 76*, 1 (2008), 53–69. 4

[SSN07] SAXENA A., SUN M., NG A. Y.: Learning 3-d scene structure from a single still image. In *IEEE Int. Conf. Comput. Vis.* (2007), IEEE, pp. 1–8. 6

[Tan08] TAN R.: Visibility in bad weather from a single image. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* (2008), pp. 1–8. 1, 2, 4

[TYW14] TANG K., YANG J., WANG J.: Investigating haze-relevant features in a learning framework for image dehazing. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* (2014). 1, 2, 3, 4, 5

[YXJ13] YAN Q., XU L., JIA J.: Dense scattering layer removal. In *SIGGRAPH Asia Technical Briefs* (2013), ACM, p. 14. 1