[[1]](#footnote-1)

A Review on Image Haze Removal Using Dark Channel Prior

Zhuowen Lin, 11510818, *Information Engineering, SUSTech*

*Abstract*—The quality of images captured by digital camera sensors can be degraded by a lot of reasons, the haze present in the atmosphere is one of which. The removal of haze, called dehazing, is typically performed upon the basis of the atmospheric light degradation model. To accomplish the task of dehazing, a statistical knowledge called dark channel prior (DCP) was proposed and later received different improvements from different research. The DCP is derived from the characteristic of outdoor haze free images that the intensity value of at least on color channel within a local image window is close to zero. Based on the DCP, the removal of haze can be finished in five steps: dark channel construction, atmospheric light estimation, transmission map construction, transmission map refinement, and image reconstruction. The five steps of dehazing not only enable us to implement them step-by-step in MATLAB, but also give us a chance to cast light on the comparison between different methods proposed by different research in each step. This review of image haze removal using dark channel prior will help readers understand the implementations and evaluations of dehazing methods based on DCP.

*Index Terms*—dehazing, image restoration

# INTRODUCTION

D

UE to the absorption and scattering by atmospheric particles in haze, outdoor images have poor quality under hazy weather. Poor quality images with low visibility bring negative effects on photography appreciation as well as computer vision applications for outdoor environments, such as object detection in satellites’ remote sensing. Haze removal, which is also named as dehazing, is considered as a critical process since clear, haze-free images can not only be visually favorable, but also be an important factor that help to improve the performance of computer image processing techniques.

Methods of dehazing before the unveiling of dark channel prior (DCP) has required multiple images to perform dehazing. For example, the polarization-based method in [1] use the polarization property of light scattering to restore the information of depth of field from multiple images taken with different degrees of polarization. However, as mentioned in the same paper [1], this method requiring multiple reference images needs a special digital image sensor to accomplish, which is not practical in daily use. This made researchers interested in finding dehazing methods relying on only single image. Most of the single-image-based methods make use of the characteristics of outdoor haze free images. For example, a method proposed in [2] takes into account the characteristic that a haze-free image has a higher contrast than a hazy image. By maximizing the local contrast of the hazy image, it can remove a certain degree of haze but introducing significant block artifacts through the discontinuities of depth of field. For another example in [3], a method infers the medium transmission by estimating the albedo of the scene. But this method builds upon an assumption that the transmission map and surface shading are locally uncorrected, which does not hold when the haze is dense.

In 2010, Doctor He proposed a paper [4] introducing a prior knowledge called dark channel prior (DCP). The DCP is based on the property of dark channel pixel, which has a very low intensity within at least one of RGB channels, except for the sky region. The DCP-based haze removal method consists of five major steps: dark channel construction, atmospheric light estimation, transmission map construction, transmission map refinement, and image reconstruction. Because DCP based method is straight forward in mathematics and easy to implement, it achieves the attentions of later studies, many of which have proposed enhancement methods of DCP.

The rest sections of this review paper are organized as follows. In Section II, the original DCP-based dehazing method proposes in [4] will be introduced. In Section III, several other adjustment methods on DCP-based one proposed by different papers will be discussed. Section IV discusses the performance evaluation of methods in Section III. Section V concludes the paper.

# Dark Channel Prior Image Dehazing Technique

In this section, the image dehazing technique based on dark channel prior (DCP) will be introduced.

## Degradation model

According to previous studies about atmospheric particles and light in [5], a hazy image can be mathematically modeled as follows:

*(1 )*

where *x* represents the image coordinates; *I* is the hazy image captured by our camera; *J* is the haze-free image or the actual object that the camera is going to shoot; *A* is the global atmospheric light; *t(x)* is the transmission map, or depth map, since it is responsible to record the information about the depth of field of a particular scene.

Since the goal of image dehazing is to recover *J* from *I*, once *A* and *t* are estimated from *I*, *J* can be arithmetically obtained as:

(2)

However, the estimation of *t* is non-trivial because *t* varies across the spatial domain according to the scene depth, the number of unknowns is equivalent to the number of image pixels. Therefore, a direct estimation of *t* from *I* without any help from prior knowledge is very difficult.

## Dark channel prior (DCP)

In [4], Doctor Kaiming He and his team did an empirical and statistical investigation of the characteristic of haze-free images. They found that there are dark pixels whose intensity values are very close to zero for at least one color channel within an local image patch. Based on this observation, they defined the dark channel as follows:

(3)

where is the intensity of the pixel in one of RGB channels and is the local patch centered at pixel x. The definition of dark channel shows that it is simply a 2-D minimum filter imposed on every local patch.

From more than 5000 outdoor haze-free images, it is shown in [4] that about 75% of the pixels in the dark channels have zero values and 90% of the pixels have values below 35. Based on this observation, they proposed a mathematical approximation:

(4)

This approximation is called dark channel prior (DCP).

Several examples of dark channels of outdoor haze free images are shown in Fig.1.

插入一张图片！

On the contrary, the pixels in dark channels of hazy images have intensities far above 0:

(5)

The existence of haze increases the intensity values in dark channel, which means that the intensities of pixels in dark channel of hazy image can act as an important clue to reveal the density of haze.

## Atmospheric light estimation

After getting dark channel image according to its definition, [4] mentions a method to estimate the atmospheric light *A*. The top 0.1% of the brightest pixels in the dark channel is first selected as candidates, and then the color with the highest intensity value among the candidate pixels is then used as the value for *A*. The whole process in shown in Fig.2.

插入一张图片！

If we ignore the first step of selecting candidate pixels, then the intensity values of white objects instead of haze regions in the original images may be wrongly selected as atmospheric light *A*, which would also be shown in Fig.2.

## Transmission map estimation

Having dark channel as well atmospheric light, the transmission map can be estimated.

The degradation model has been introduced as Equation (1).

If we divide both sides of the equation by atmospheric light *A*, then we can have the minimum intensity in the local patch of each color channel:

*(6)*

Here, the transmission map within each local patch is assumed to be constant as in [4].

Then we can introduce the definition of dark channel by further imposing a 2-D minimum filter concerning RGB channel in every local patch:

(7)

Because of the dark channel prior (DCP) approximation:

(8)

the transmission map can be written as:

(9)

In practical use, we often introduce a parameter ranging from 0 to 1into Equation (9) to maintain a certain degree of haze in the final result. This is because haze or fog is one of the important sources of sense of depth of field for our human eyes. If we eliminate all the haze, then the image would be too artificial.

(10)

This is the basis of transmission map calculation from dark channel and atmospheric light *A*.

## Transmission map refinement

If the transmission map from section D is directly used to reconstruct the dehazed image, one of the example are shown in Fig.3. When using un-refined transmission map, most of the haze in the original image can be removed, while the obvious block artifacts around the edges of objects are also introduced into the dehazed image. Therefore, it is of necessity to do refinement of transmission map.



Fig. 3. Dehazed image when using un-refined transmission map.

In [4], the refinement method is called soft matting, which is a method originally used in Laplacian image matting [6].

As introduced in [4] and [6], if we denote the refined transmission map as and the un-refined transmission map as . Rewriting and in their vector forms and then the refinement of transmission map can be done by minimizing the following cost function:

Here, the matrix L is called the matting Laplacian matrix introduced in [6] as: 打那个L的公式！

The optimal can be obtained by solving the following sparse linear system function:

where U is an identity matrix of the same size as L.

However, the soft matting method, in particular the calculation of matting Laplacian matrix L, consumes a large amount of computer memory, making it impractical for images in large size.

In one of his later proposed papers [7], Doctor Kaiming He solved this problem by replacing soft matting with a new method called guided filtering.

The guided filter computes the filtering output by considering the content of a guidance image, which can be the input image itself or another different image. It can transfer the structures of the guidance image to the filtering output. As introduced in [7], the basic assumption of guided filtering is a local linear model between the guidance image and the filtering output. If *q* is denoted as a linear transform of *I*, then the guided filtering process is defined as: 打那3个公式！

The guided filtering process is also illustrated in Fig. 4. 截文章的图！

In [7], the authors also give the algorithm to do guided filtering for image, as shown below. 截图或者打那个算法！

After the refinement by soft matting or guided filtering, the transmission map would be shown in Fig. 5. Compared with the unrefined one, the refined transmission map has the basic shape and the depth of field of the original image.



Fig. 5. Refined transmission map.

## Image reconstruction

After having atmospheric light *A* and refined transmission map *t*(x), the reconstruction of dehazed image can be done easily according Equation (2).

The dehazing results are shown below. The dehazing algorithm successfully accomplish its task, since most of the haze is removed and a certain degree of haze is maintained intentionally from the consideration of sense of depth of field of our human eyes.



Fig. 6. Demonstration of dehazing result

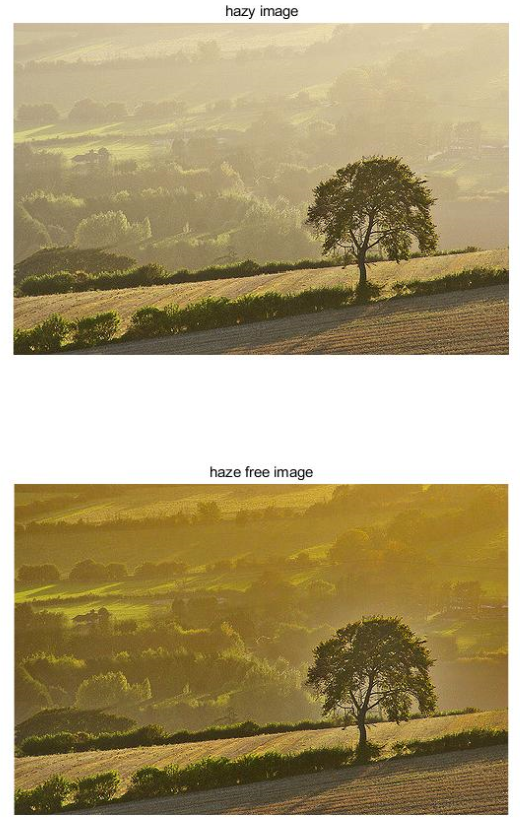


Fig. 7. Demonstration of dehazing result

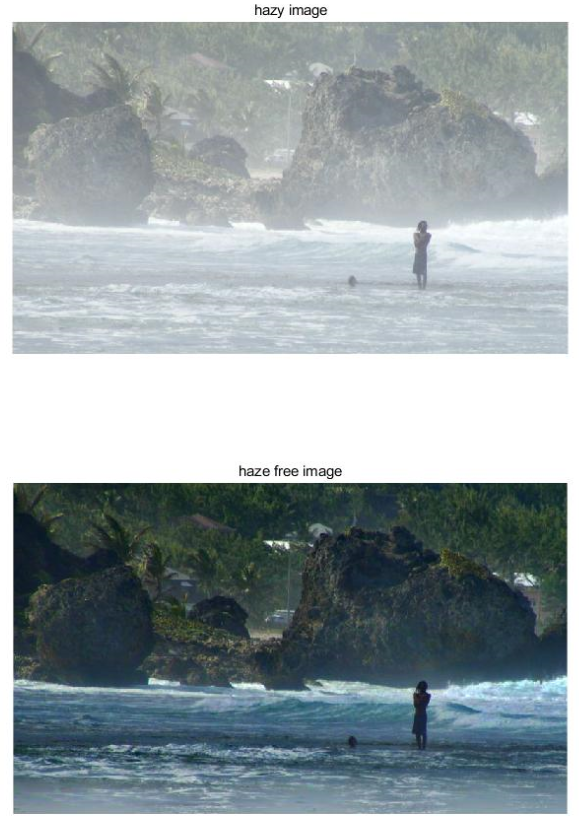


Fig. 8. Demonstration of dehazing result

## Discrete cosine transform (DCT)

There are mainly two reasons for taking DCT: one is that DCT can decorrelates the overlapping filter bank energies, which means that diagonal covariance matrices can be used to model the features in a Hidden Markov Model (HMM) classifier. Another one is that because DCT has cosines with half-integer numbers of cycles, it can compactly fit the shape of power spectrum energies with few DCT coefficients, which is better than DFT. For a particular frame in the speech file “s5.wav”, I write a MATLAB program to test how many DCT coefficients are sufficient to represent over 99% of the energy. The result shows that only 84 out of totally 208 DCT coefficients are sufficient to represent over 99% of the energy. If we use these 84 coefficients to reconstruct the speech signal in this frame, comparing with the original frame, the result would be shown in Fig. 4.

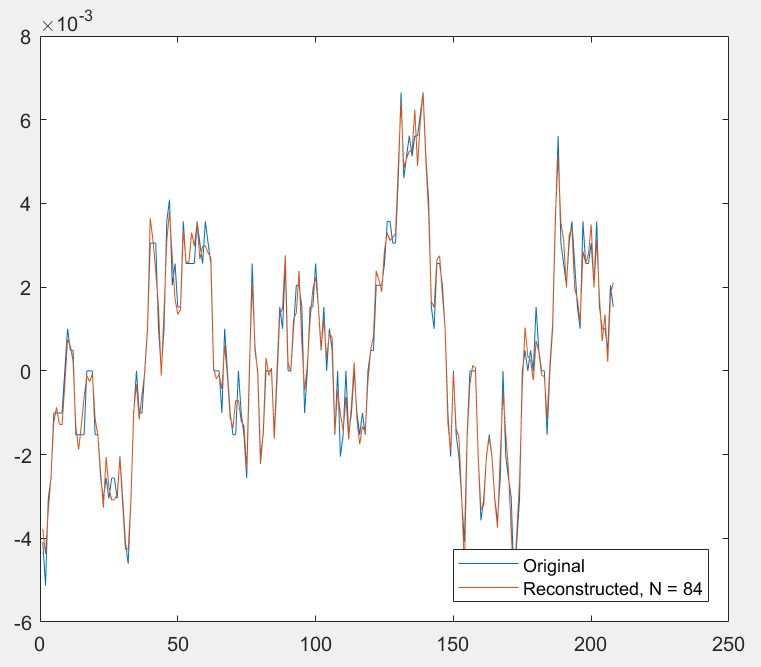


Fig. 4. Original speech frame and reconstructed speech frame from 84 DCT coefficients.

## Coefficient discarding

In fact, 84 coefficients are redundant for MFCC. Only 12 DCT coefficients are kept since the higher DCT coefficients represent fast changes in the filter bank energies and it turns out in [5] that these fast changes are not useful in speech recognition. By discarding coefficients can further simplify our parametric representation.

## Delta coefficient

The 12 MFCC feature coefficients describe only the power spectral envelope of a single frame, but the speech would also have information lying in dynamics, such as the changing trend of the MFCC. Therefore, it would be helpful to append delta coefficients after MFCC.

The delta coefficients are defined as:

where is a delta coefficient, t is the current frame number, is the MFCC in frame n. By introducing MFCC in previous and later frames into delta coefficients, the dynamic changing trend of MFCC has been preserved.

# Complex Linear Projection

In [9] proposed by Google Inc., the authors claimed a disadvantage of MFCC, saying MFCC separates the perceptually motivated filters from the acoustic model, which is not always the best choice in statistical modeling frameworks such as automatic speech recognition (ASR), where the end goal is word error rate minimization.

In order to overcome this disadvantage, the authors introduced a neural-network-based technique called Complex Linear Projection (CLP), which performs both filtering layer and pooling layer in neural network in frequency domain and produces set of ASR features which can be fed to the backend neural network acoustic model. By doing this, the CLP model can be jointly optimized with the acoustic model.

## Comparison between CLP and MFCC

Introduced in [9], the output of the filter in CLP model is:

.

where is a complex matrix called projection matrix in . is the DFT of input signal.

The CLP model is different from MFCC in terms of their pooling operation: MFCC uses pooling while CLP uses a simple summation pooling. The MFCC filter bank operates on features while CLP model operates on . This means that the phase information in is removed in the MFCC while it is preserved in CLP model. There is a long-time debate about the improvement of phase for single microphone speech recognition. However, in multi-channel recognition, it is agreed by [10] that phase information is necessary, because it can preserve the relative delay of the speech signal at each microphone channel.

## Experiment results comparing CLP and MFCC

In one of the experiments in [9], the performance of CLP is evaluated in terms of the effectiveness of CLP model to achieve state-of-the-art performance using MFCC. The results are shown in Table II and Table III.

Table II presents the baseline WERs for MFCC model across three different window sizes. Typically, a window size of 25 msec with a 10 msec shift is used in ASR. As the results show, longer window contains more temporal information as well as localization of multi-channel processing which result in WER improvement over shorter windows.

The baseline CLP model results are shown in Table III. The CLP model performance is in the same level as MFCC model for single channel but yields a gain of about 4% over the 2-channel baseline models in Table II.

TABLE II

Word Error Rate for the Baseline Models of MFCC

|  |  |  |
| --- | --- | --- |
| Model | 1-Channel | 2-Channel |
| 25 msec | 23.4 | 21.8 |
| *32 msec* | 22.8 | 21.3 |
| *64 msec* | 21.8 | 20.7 |

TABLE III

Word Error Rate for the Baseline Models of CLP

|  |  |  |
| --- | --- | --- |
| Model | 1-Channel | 2-Channel |
| 25 msec | 23.2 | 21.5 |
| *32 msec* | 22.8 | 20.9 |
| *64 msec* | 22.0 | 20.5 |

To sum up, the Complex Linear Prediction (CLP) model achieves superior performance compared to MFCC by preserving all the information in the signal, including the time delay or phase information from multiple channels in a microphone array, which makes it appropriate to automatically learn the optimal feature extraction parameters in the multi-channel setting at the help of neural network.

# Conclusions

This paper cast light on the development stages of mel-frequency cepstrum coefficients (MFCC). The cepstrum-based linear prediction cepstrum coefficients (LPCC) is firstly introduced. Its calculation is based on smoothed auto-regressive power spectrum or recursive formula from LPC. However, LPCC is inferior to MFCC in the comparison between parametric representations for speech recognition regardless of the frame separation, type of testing or speaker. MFCC is introduced secondly. The better performances of MFCC are largely due to its perceptual-based mel-frequency scale, which can better model the human hearing. The computation of MFCC is divided into six steps. The paper reviewed the mathematical formulas as well as MATLAB implementations for these six steps. A technique yielding superior performances over MFCC called complex linear projection (CLP) is included thirdly. By preserving the phase angle information in CLP, it can have a smaller word error rate than MFCC in multi-channel speech recognition.

TABLE I

Units for Magnetic Properties

|  |  |  |
| --- | --- | --- |
| Symbol | Quantity | Conversion from Gaussian and  CGS EMU to SI a |
| Φ | magnetic flux | 1 Mx → 10−8 Wb = 10−8 V·s |
| *B* | magnetic flux density,  magnetic induction | 1 G → 10−4 T = 10−4 Wb/m2 |
| *H* | magnetic field strength | 1 Oe → 103/(4π) A/m |
| *m* | magnetic moment | 1 erg/G = 1 emu  → 10−3 A·m2 = 10−3 J/T |
| *M* | magnetization | 1 erg/(G·cm3) = 1 emu/cm3  → 103 A/m |
| 4π*M* | magnetization | 1 G → 103/(4π) A/m |
| σ | specific magnetization | 1 erg/(G·g) = 1 emu/g → 1 A·m2/kg |
| *j* | magnetic dipole  moment | 1 erg/G = 1 emu  → 4π × 10−10 Wb·m |
| *J* | magnetic polarization | 1 erg/(G·cm3) = 1 emu/cm3  → 4π × 10−4 T |
| χ*,* κ | susceptibility | 1 → 4π |
| χρ | mass susceptibility | 1 cm3/g → 4π × 10−3 m3/kg |
| μ | permeability | 1 → 4π × 10−7 H/m  = 4π × 10−7 Wb/(A·m) |
| μr | relative permeability | μ → μr |
| *w, W* | energy density | 1 erg/cm3 → 10−1 J/m3 |
| *N, D* | demagnetizing factor | 1 → 1/(4π) |

Vertical lines are optional in tables. Statements that serve as captions for the entire table do not need footnote letters.

aGaussian units are the same as cg emu for magnetostatics; Mx = maxwell, G = gauss, Oe = oersted; Wb = weber, V = volt, s = second, T = tesla, m = meter, A = ampere, J = joule, kg = kilogram, H = henry.

References

1. YY Schechnner, SG Narasimhan, SK Nayar, “Polarization-based vision through haze,” *Appl. Optics* 42(3), 511-525 (2003)
2. RT Tan, “Visibility in Bad Weather from a Single Image,” in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (CVPR, Anchorage, 2008), pp. 1-8
3. R Fattal, “Single Image Dehazing,” *ACM Trans. Graph*. 72(3), 72:1-72:9 (2008)
4. K He, J Sun, X Tang, “Single Image Haze Removal using Dark Channel Prior,” in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, No. 12, 2011
5. N Hautiere, JP Tarel, J Lavenant, D Aubert, “Automatic Fog Detection and Estimation of Visibility Distance Through Use of Onboard Camera,” Mach. *Vis. Appl*. 17(1), 8-20 (2006)
6. A. Levin, D. Lischinski, and Y. Weiss, “A Closed Form Solution to Natural Image Matting,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 61-68, 2006
7. K He, J Sun, X Tang, “Guided Image Filtering,” in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, 2013
8. P. Mermelstein, “Distance measures for speech recognition, psychological and instrumental,” in *Pattern Recognition and Artificial Intelligence*, C. H. Chen, Ed. New York: Academic, 1976, pp. 374-388.
9. L. R. Rabiner and R. W. Schafer, “The Cepstrum and Homomorphic Speech Processing,” in *Theory and Applications of Digital Speech Processing*, 1st ed. Upper Saddle River, NJ, United States: Prentice Hall, 2011, ch. 8, sec. 4, pp. 429-440
10. X. Huang, A. Acero and H. Hon, “Cepstrum of Speech Signals,” in *Spoken Language Processing*, 1st ed, Upper Saddle River, NJ, United States: Prentice Hall, 2001, ch. 6, sec. 4, pp.312-314
11. S. B. Davis and P. Mermelstein, “Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentecnces,” *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 28, no. 4, pp. 357-366, 1980
12. X. Huang, A. Acero and H. Hon, “Speech Perception,” in *Spoken Language Processing*, 1st ed, Upper Saddle River, NJ, United States: Prentice Hall, 2001, ch. 2, sec. 1, pp.33-34
13. Steven, S.S. and J. Volkman, “The Relation of Pitch to Frequency”, *Journal of Psychology*, 1940, 53, pp. 329
14. X. Huang, A. Acero and H. Hon, “Mel Frequency Cepstrum,” in *Spoken Language Processing*, 1st ed, Upper Saddle River, NJ, United States: Prentice Hall, 2001, ch. 6, sec. 5, pp.316-318
15. E. Variani, T.N. Sainath, I. Shafran, and M. Bacchiani, “Complex Linear Projection (CLP): A Discriminative Approach to Joint Feature Extraction and Acoustic Modeling,” in INTERSPEECH, San Francisco, CA, USA, 2016
16. T. N. Sainath, R. J. Weiss, K. W. Wilson, A. Narayanan, M. Bacchiani, and A. Senior, “Speaker Localization and Microphone Spacing Invariant Acoustic Modeling from Raw Multichannel Waveforms,” in *Proc. ASRU*, 2015.

1. Zhuowen Lin is with the Information Engineering major, Southern University of Science and Technology, Shenzhen, Guangdong, China (e-mail: 11510818@mail.sustc.edu.cn). [↑](#footnote-ref-1)