Inferring the level of visibility from hazy images

Alexander A.S. Gunawan*

Mathematics Department,
Bina Nusantara University,
KH Syahdan 9, Jakarta 11530, Indonesia
Email: aagung@binus.edu
*Corresponding author

Heri Prasetyo

Informatics Department,
Sebelas Maret University,
Ir. Sutami No. 36A, Surakarta,
Jawa Tengah, Indonesia
Email: heri.inf.its.02@gmail.com

Indah Werdiningsih

Mathematics Department,
Airlangga University,
Muljorejo Kampus C,
Surabaya, Indonesia
Email: Indah_werdiningsih@yahoo.co.id

Janson Hendryli

Informatics Engineering Department,
Tarumanagara University,
Letjen S. Parman No. 1, Jakarta, Indonesia
Email: jansonh@fti.untar.ac.id

Abstract: In our research, we would like to exploit crowdsourced photos from social media to create low-cost fire disaster sensors. The main problem is to analyse how hazy the environment looks like. Therefore, we provide a brief survey of methods dealing with visibility level of hazy images. The methods are divided into two categories: single-image approach and learning-based approach. The survey begins with discussing single image approach. This approach is represented by visibility metric based on contrast-to-noise ratio (CNR) and similarity index between hazy image and its dehazing image. This is followed by a survey of learning-based approach using two contrast approaches that is: 1) based on theoretical foundation of transmission light, combining with the depth image using new deep learning method; 2) based on black-box method by employing convolutional neural networks (CNN) on hazy images.

Keywords: hazy image; visibility level; single image approach; learning-based approach; social media.

Reference to this paper should be made as follows: Gunawan, A.A.S., Prasetyo, H., Werdiningsih, I. and Hendryli, J. (2020) 'Inferring the level of visibility from hazy images', *Int. J. Business Intelligence and Data Mining*, Vol. 16, No. 2, pp.177–189.

Biographical notes: Alexander A.S. Gunawan received his Bachelor in Mathematics from Bandung Institute of Technology (ITB), Indonesia and MSc in Automation Engineering from Fachhochschule Darmstadt, Germany. He received his PhD from University of Indonesia in Computer Science. Currently, he is a faculty member with Mathematics Department, Bina Nusantara University, Indonesia. His research interests are robotics, computer vision and machine learning.

Heri Prasetyo received the BS degree from Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia, in 2006, and the MS and PhD degrees from National Taiwan University of Science and Technology (NTUST), Taipei, Taiwan, in 2009, and 2015, respectively. Currenlty, he is a faculty member with the Department of Informatics, Universitas Sebelas Maret (UNS), Surakarta, Indonesia. His research interests include image processing, multimedia signal processing, machine learning, and pattern recognition.

Indah Werdiningsih received his SSi in Mathematics from Institut Teknologi Sepuluh Nopember, Indonesia and her MKom in Informatics Engineering from Institut Teknologi Sepuluh Nopember, Indonesia. Currently, she is a researcher in Faculty of Science and Technology at Universitas Airlangga, Indonesia. Her primary research interests include data mining, fuzzy systems, and expert systems.

Janson Hendryli received his SKom in Informatics from Tarumanagara University, Indonesia and his MKom in Computer Science from Universitas Indonesia. Currently, he is a lecturer at Tarumanagara University, Indonesia. His primary research interests include deep learning and computer vision.

1 Introduction

Haze image occur because bad weather conditions such as haziness, mist, foggy and smoky. The image quality of outdoor scene in the fog and haze weather condition is usually deteriorated by the scattering of a light before reaching the camera due to these large quantities of suspended particles (e.g., fog, haze, smoke, impurities) present in the atmosphere (Saggu and Singh, 2015). The presence of haze in the atmosphere degrades the quality of images captured by visible camera sensors. The removal of haze, called dehazing, is typically performed under the physical degradation model, which requires a solution of an ill-posed inverse problem (He et al., 2009). Therefore, improving the technique of image haze removal will benefit many image understanding and computer vision applications such as aerial imagery (Woodell et al., 2006).

In our research, we would like to exploit crowd-sourced photos from social media to create a low-cost alternative to fire disaster sensors. The main problem in here is to analyse how hazy the environment looks like from images. Therefore we provide a brief survey of methods dealing with visibility level of hazy images by dividing the methods into two categories. First method is only based on single-image, that visibility metric (VM) and similarity index. VM produces a metric based on contrast-to-noise ratio (CNR). This metric is based on the computation of computation of the standard-deviation image and can be used to judge which dehaze method is better than another one, since it provides a quantitative metric for haze images. The VM is proposed for judging which dehaze method is better (Zhengguo, 2012). The similarity index is a method for predicting the perceived quality of digital television and cinematic pictures, as well as other kinds of digital images and videos. This index, called as structural similarity (SSIM) (Wang et al., 2004), is used for measuring the similarity between two images. In here, we compare the original hazy image to dehazing image using SSIM to estimate the visibility level of hazy image.

Second approach is based on learning paradigm. Haze removal is actually a difficult task because fog depends on the unknown scene depth map information. Based physical observation and theory, fog effect is the result of light transmission and distance between camera and object. Hence removal of fog requires the estimation of air-light map (Saggu and Singh, 2015). Furthermore, He et al. (2009) found that there are dark pixels whose intensity values are very close to zero for at least one-colour channel within an image patch, called as dark channel. Approximation to zero for the pixel value of the dark channel is called the dark channel prior (DCP) (Lee et al., 2015). The DCP is based on the statistics of haze-free outdoor images. Combining a haze imaging model and a soft matting interpolation method, we can recover a hi-quality haze-free image. By using a transmission matrix generated from DCP algorithm (He et al., 2009), and depth map from new deep convolutional neural fields (DCNF) method (Liu et al., 2015), haze level score can be computed by combining the transmission matrix and depth map. The depth map, in here, is estimated depths from single monocular images (Liu et al., 2015). We consider the transmission matrix as the perceived depth of hazy photos, which is a combination of actual depth and haze effects. Therefore, by ruling out the actual depth factor, we can isolate the haze effects from the transmission matrix, which is used to estimate the haze level (Li et al., 2015). Recently, there is evidence that the black-box approach using deep convolutional neural networks (CNN) are setting new records for various vision applications. To consider this black-box approach, we also create CNN based on hazy images and their visual evaluation.

We propose VM, similarity index, theoretical approach based on transmission and depth map, and also black-box approach based on CNN. The experiment will compare the results these four approaches.

2 Single image approach

In single image-based approach, we infer the haze visibility level only by one image through image processing techniques, such as VM and image similarity index which explain in detail in next subsection.

2.1 Visibility metric

The VM is based on the computation of the standard-deviation image and can be used to judge which dehaze method is better than another one, since it provides a quantitative metric for haze images. This VM is calculated by using CNR (Wang et al., 2004) of noise image estimated by Gaussian kernel. CNR is a measure used to determine image quality. CNR is similar to the metric signal-to-noise ratio (SNR), but subtracts off a term before taking the ratio. This is important when there is a significant bias in an image, for example in hazy image which the features of the image are washed out by the haze. Thus, this image may have a high SNR metric, but will have a low CNR metric. We conduct experiments based on the VM from Zhengguo (2012).

2.2 SSIM index

The SSIM index (Wang et al., 2004) is a method for predicting the perceived quality of digital television and cinematic pictures, as well as other kinds of digital images and videos. It was first developed in the Laboratory for Image and Video Engineering (LIVE) at the University of Texas at Austin and in subsequent collaboration with New York University. SSIM is used for measuring the similarity between two images. The SSIM index is a full reference metric; in other words, the measurement or prediction of image quality is based on an initial uncompressed or distortion-free image as reference. SSIM is designed to improve on traditional methods such as peak signal-to-noise ratio (PSNR) and mean squared error (MSE), which have proven to be inconsistent with human visual perception.

To estimate the visibility level of hazy image, we compare the original hazy image to dehazing image using SSIM. In order to utilise SSIM, one renowned image dehazing method, called as multi-scale fusion algorithm (Ancuti and Ancuti, 2013), is used to do haze removal

2.2.1 Single image dehazing based on multi-scale fusion

Haze is an atmospheric phenomenon that significantly degrades the visibility of outdoor scenes. This is mainly due to the atmosphere particles that absorb and scatter the light. For our experiment, we used multi-scale fusion algorithm for removing have in an image. This fusion-based strategy works by applying a white balance and a contrast enhancing to two original hazy image inputs. To blend effectively the information of the derived inputs to preserve the regions with good visibility, we filter their important features by computing three measures (weight maps): luminance, chromaticity, and saliency (see Figure 1). To minimise artefacts introduced by the weight maps, it is designed in a multi-scale approach, using a Laplacian pyramid representation. The implementation of multi-scale fusion algorithm is appropriate for real-time applications (Sreekuttan, 2016).

The below procedure (Padole and Khare, 2015) is foundation of the multi-scale fusion algorithm, that is:

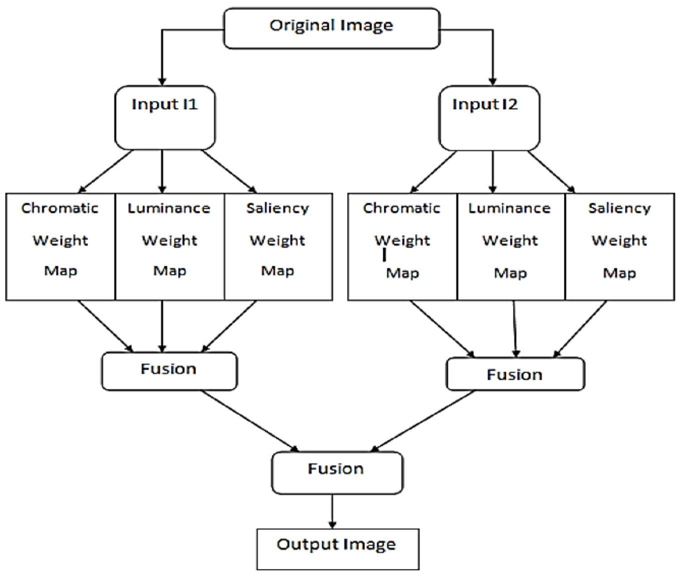
1. Derive two input images from the original input with the aim of recovering the visibility for every region of the scene in at least one of them:
   a. first input will be obtained by applying white balancing
   b. second input will be obtained by applying contrast enhancement technique.
2. Compute three weight maps such as luminance, chromaticity and saliency and weight the derived inputs by three normalised weight maps.

3 Apply multi-scale fusion, utilising Laplacian pyramid delegation of inputs blended along with Gaussian pyramids of normalised weights to obtain the haze free image.

4 Apply unsharp masking method (USM) for image dehazing on original hazy input image to obtain the haze free image.

5 Compare the results of single image dehazing using multi-scale fusion method with USM of single image dehazing to prove the efficiency of USM.

Figure 1 Multi-scale fusion algorithm



3 Learning-based approach

In contrast to single image approach, learning-based approach envoy a model learned from many images to help inferring the haze visibility. In here, we employ two contrast approaches, that is:

1 based on theoretical foundation of transmission light, combining with the depth image using new deep learning method

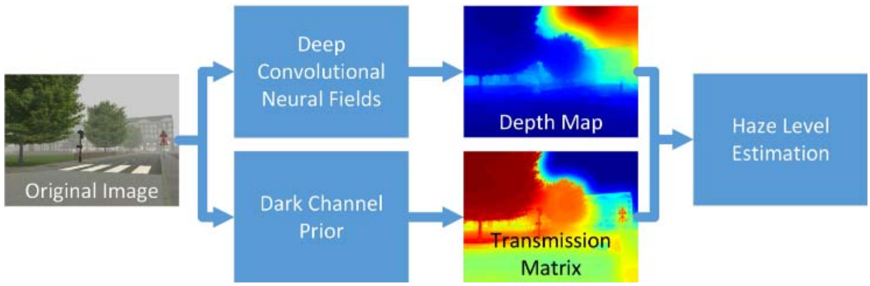
2 based on black-box method by employing deep learning on hazy images and their visual evaluation.

Therefore, there are two learning-based approaches that we use in this research: DCNFs to learn depth from images and CNN to classify haze level into two classes of heavy and light haze.

3.1 Depth map and transmission matrix

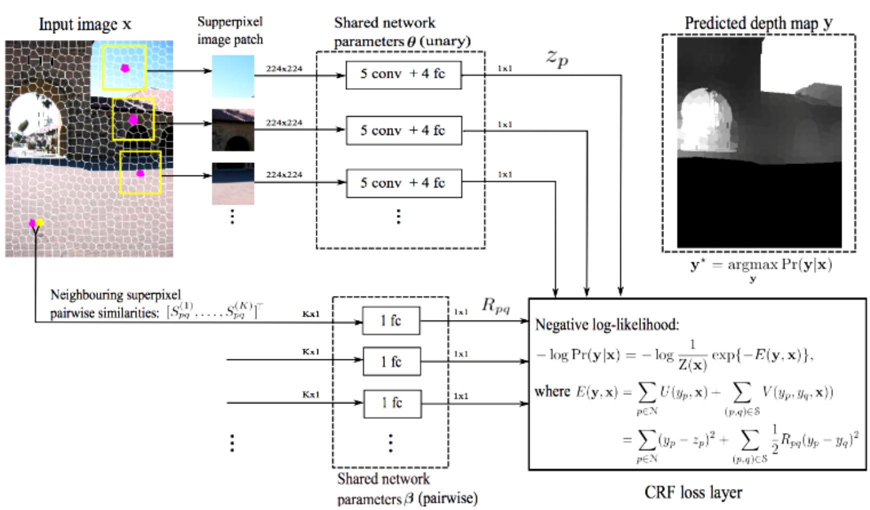
The basic idea of this approach is that the haze visibility can be inferred by computing the depth of the images and the transmission matrix generated from a haze removal algorithm. In paper, Li et al. (2015) proposed a framework to estimate haze level from a photo by using the DCP (He et al., 2009) to estimate the transmission matrix and the DCNF (Liu et al., 2015) to estimate the depth map. By combining this information, they select from a combination of transformation and pooling functions to estimate the haze level. Figure 2 shows the proposed framework of (Li et al., 2015).

Figure 2 The proposed framework of Li et al. (2015) to estimate the haze level from photos (see online version for colours)



The DCNF is proposed by (Liu et al., 2015) to estimate depth information given a single monocular image. The model combines the strength of CNN and conditional random fields (CNF) to predict the depths. Experimental results from Liu et al. (2015) show improved accuracy on various dataset and other baseline methods. Figure 3 shows the illustration of the DCNF model. For the DCNF, we use the source code and learned model provided by the authors (https://bitbucket.org/fayao/dcnf-fcsp).

Figure 3 DCNF model (see online version for colours)



Source: Liu et al. (2015)

Our approach is quite similar to Li et al. (2015), but we choose to use following equation to estimate the haze level from the estimated depth map and transmission matrix:

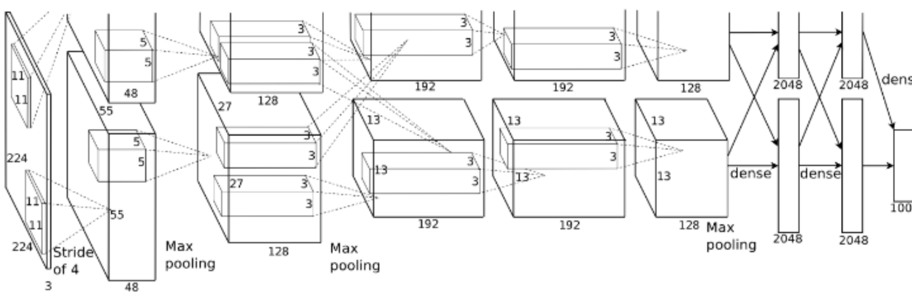
$$k = \text{median}\left(\frac{\log t(x)}{d(x)}\right) \tag{1}$$

where k denotes the haze level, $t(x)$ is the transmission matrix, and $d(x)$ is the depth map. Our choice of transformation and pooling function are due to the time and computation constraint.

3.2 Convolutional neural network

Nowadays, there is evidence that the black-box approach using deep CNNs are setting new records for various vision applications. To consider this black-box approach, we also experimented on CNNs to find the possibility using black-box approach in calculating the visibility index. CNN has been arguably the best image classifier since Krizhevsky, et al. won the ImageNet object classification in 2012 using deep CNN (Krizhevsky et al., 2012). CNN is basically a feed-forward neural network but with a convolutional layer for the purpose of learning the best representation of the images. Early implementation of CNN, widely known as LeNet-5 (LeCun et al., 1989), applied by several banks to recognise handwritten digits on cheques.

Figure 4 Architecture of AlexNet

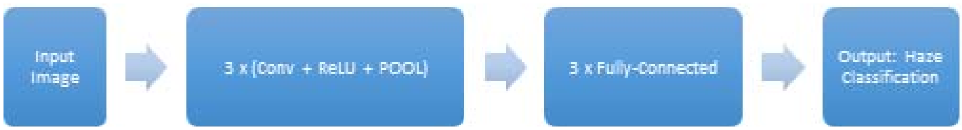


Source: Krizhevsky et al. (2012)

Figure 4 shows the architecture of Krizhevsky's deep CNN, widely known as AlexNet, which won the ImageNet challenge on 2012. The networks consist of five convolutional layers, followed by three layers of fully connected networks. On the first, second, and fifth convolutional layers, a max-pooling layer is applied to summarise the outputs of adjacent neurons in the same layer (Krizhevsky et al., 2012). In addition to that, it applies the ReLU activation functions to the output of every convolutional layer. Finally, the last layer is a softmax which produces a distribution over the 1,000 class labels. Moreover, a dropout regularisation is used in the first two fully connected layers to reduce overfitting. The AlexNet model is trained using stochastic gradient descent (SGD) with a batch size of 128, momentum 0.9, and weight decay of 0.0005.

We implement the above CNN architecture to classify hazy images into two classes of heavy haze and light haze using Keras and Python. Unfortunately, the CNN classification gave discrete value outputs, so it does not produce any value to represent visibility level. Our network consists of six layers, of which three layers are the convolutional layers and three layers are the fully connected. We use max pooling of size 2×2, ReLU activation on each convolutional layer, and a dropout regularisation on the first and second fully connected layers. The output of the last layer is fed to a sigmoid function. We also train the network using SGD with a batch size of 32 and 100 epoch. Furthermore, image augmentations are applied to each training images, which rotate, translate, rescale, zoom, and horizontally flip the images to reduce overfitting. The augmentations are performed using Keras ImageDataGenerator library. Figure 5 illustrates the configuration of our CNN architecture. Furthermore, we evaluate the performance of CNN architecture by comparing to well-known classifier support vector machine (SVM) based on LibSVM and PIL library in Python.

Figure 5 Architecture of our CNN architecture to classify haze level (see online version for colours)



4 Experiment results

4.1 Datasets

For CNN, we manually classify the data from social media provided by Pulse Lab Jakarta into two classes of haze level: heavy and light. We obtain 300 images for training (191 images for heavy haze and 109 images for light haze) and 57 images for testing. Furthermore, we test the CNN model with five additional images with higher resolution. Since the social media images where the dataset comes from are heavily filtered, i.e., Instagram filters, we need to test the robustness of the model using non-social media images. These additional images are retrieved from simple Google search of haze images.

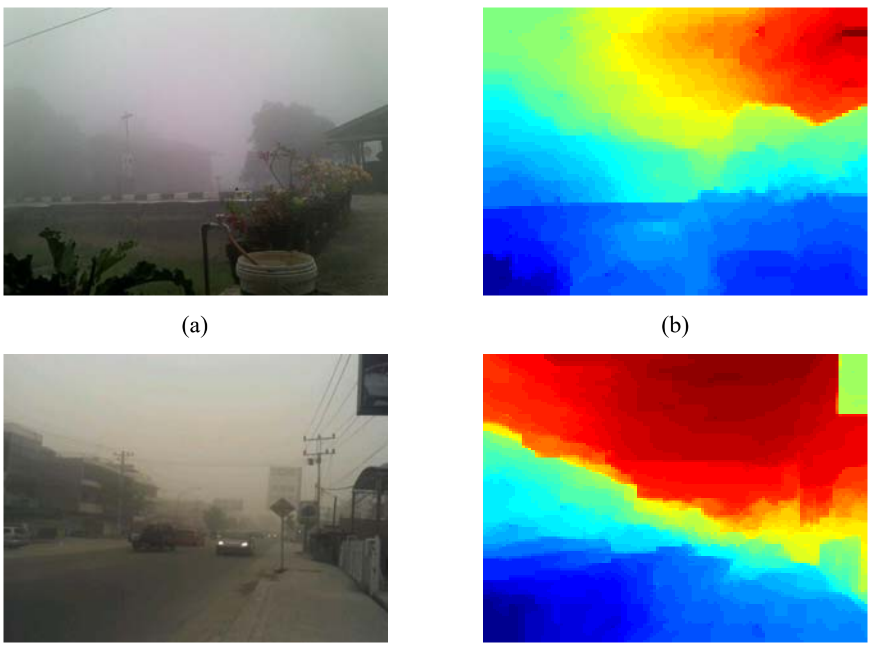
4.2 Experiments on single image approach

VM estimates how far visibility level based on single image. Thus, lower VM should represent thicker haze. The result of VM on Pulse Lab hazy images dataset can be seen in Figure 7. VM is only suitable to compare the same outdoor images with different haze level, so VM is not too reliable in estimating visibility level of many unrelated outdoor images taken from social media. On the other hand, SSIM index is based on comparison of original hazy image and its dehazing image. Therefore, SSIM is depended on the dehazing image algorithm. For our experiment, we choose recent multi-scale fusion algorithm (Ancuti and Ancuti, 2013). Because SSIM essentially measure similarity distance, then larger SSIM should represent thicker haze. The result of SSIM on Pulse Lab hazy images dataset can be seen in Figure 7.

4.3 Experiments on learning-based approach

The result of running DCNF on Pulse Lab hazy images dataset can be seen in Figure 6 below. The heat map should be corresponding to the distance between the objects on the images and the camera. As we can see from the figure, the depth map has troubled inferring the depth of the sky and heavy haze-covered objects. Calculating haze level k using equation (1) gives us results shown in Figure 7. The transmission matrices $t(x)$ used in this approach come from Dark Prior approach (He et al., 2009). The larger value of k should mean the haze is heavier, though there are some inaccuracies. For example, for Figure 7(c), the value of k is larger than Figure 7(d), in spite of the heavier haze level that can be seen visually.

Figure 6 Result of DCNF on Pulse Lab dataset, images pair (a)–(b) and (c)–(d) show the original images and the corresponding depth map (see online version for colours)



(a) (b) (c) (d)

Figure 8 shows the classification of the five additional high-resolution images using CNN model, trained on 300 social media images provided by Pulse Lab Jakarta. The CNN model correctly classifies the heavy haze images, illustrated in Figures 8(d) and 8(e), although Figure 8(c) can be argued as light or heavy haze. It can be concluded that the CNN specifically recognise the heavy haze images from the visibility level of objects in an image. For the Pulse Lab dataset, we get around 0.75 accuracy though the training epoch has not really converged due to the time constraints.

Figure 7 Estimated haze level SSIM, *k* using DCNF and transmission matrix (see online version for colours)

VM = 31.7657
SSIM = 0.4141
k = 0.022964
(a)

VM = 28.4074
SSIM = 0.6229
k = 0.027969
(b)

VM = 50.2235
SSIM = 0.6042
k = 0.031128
(c)

VM= 31.2852
SSIM= 0.7515
k = 0.029282
(d)

VM=41.2091
SSIM= 0.7964
k = 0.032166
(e)

VM= 21.0301
SSIM=0.5256
k = 0.04007
(f)

Note: Larger SSIM and *k* represent thicker haze, but lower VM should represent thicker haze.

Figure 8 CNN classification result of the five additional images, images (a)–(c) are classified as light haze, while images (d) and (e) are correctly classified as heavy haze (see online version for colours)

(a)

(b)

(c)

(d)

(e)

In this study, we have evaluated several methods for estimating visibility level that is: VM, SSIM, the depth map + dark channel prior (DCNF+DCP) and CNN. To make brief conclusion, we calculated Spearman correlation index like in Li et al. (2015) based on the experiment results and human expert evaluation of Pulse Lab hazy images dataset. Unfortunately, our CNN classification approach does not give any value to represent visibility level, thus the correlation index cannot be calculated.

Table 1 Spearman correlation coefficients (%) performance

VM	*SSIM*	*DCNF + DCP*	*CNN*
–0.29	-0.08	0.66	-

Furthermore, in our short time experiment, we managed to train six layers of CNN and got the 75% classification accuracy which looks so promising. In this training, the CNN model has not converged well. Yet, validation on additional five images with higher resolution gives arguably good results. The hyper-parameters of the CNN can also be fine-tuned more, i.e., by choosing better layer configuration. To compare the CNN performance, we also classified the same dataset into two categories: heavy haze and light haze by using SVM based on LibSVM and PIL library in Python. The dataset is divided into the ratio of 70:30, where 70% is for training and 30% is for testing. Overall classification accuracy of linear SVM is about 57.14%. We thought the lower accuracy of linear SVM comparing to CNN is due by SVM cannot extract the important features from raw images. The hazy image cannot be inferred just from pixel intensity, but need spatially local correlation among pixels. On the other hand, CNN layers in deep learning architecture can capture spatial locality by learning a local connectivity pattern among pixels in the image, before the fully connected layers in the end build an image classifier. The third fully connected layer has only two outputs which represent the probability of an image being heavy haze and light haze.

Additionally, mean normalisation can be used for the input images, so hopefully, the loss functions can converge more quickly and we could get the best model to predict the haze level from images. The hyper-parameters of the CNN can also be fine-tuned more, i.e., by choosing better layer configuration.

5 Conclusions

We managed to experiment on several approaches to estimate visibility level using two main methods: single-image approach and learning-based approach. Based on correlation coefficient (see Table 1), both single-image approaches: VM and SSIM obtain undesired results, that there is no significant correlation between the visibility level results and human expert evaluation on experiment dataset. However, the DCNF + DCP, which is learning based approach, gives the promising result. Its correlation when comparing to human expert evaluation reaches 66%, but there are still some inaccuracies. For example, the comparison between Figures 7(c) and 7(d), the subjective human expert evaluation said haze in Figure 7(d) is thicker than Figure 7(c). On the other hand, the DCNF+DCP inferred Figure 7(c) have thicker haze (k = 0.031128) than in image Figure 7(d) (k = 0.029282). We argue that the inaccuracy problem is due to the unavailability of ground truth depth map for Pulse Lab dataset. Consequently, the human expert just used his subjective common sense and we also cannot train the DCNF to predict the depth from the provided dataset. Another problem is the image dataset from social media contains various artificial image processing, so that the visibility level model is incorrectly estimated. For future research, the DCNF should be trained on social media images with depth information to be able to model accurately the variability of image representation. The other approaches (VM, SSIM) seem only suitable for well-behaved image dataset, and do not fit for our ill-behaved social media image dataset.

Inspired by the surprising results of deep learning, we find the possibility using black-box approach in calculating the visibility level. We constructed the ordinary CNN architecture and evaluated its performance to classify hazy images into two classes of heavy haze and light haze. The classification gives only discrete value outputs, so it does not produce any value to represent visibility level. Nevertheless, our CNN architecture gives initial promising result (about 75% accuracy) in classifying two classes of hazy images. For future research, we should create the modification of CNN architecture by adding regression layer, so that it can produce continuous value output as representation of visibility level.

Acknowledgements

We would like to thank to Pulse Lab Jakarta for introduce us to the problem discussed in this paper, as well as for provides data used in the experiments.

References

Ancuti, C.O. and Ancuti, C. (2013) 'Single image dehazing by multi-scale fusion', *IEEE Transactions on Image Processing*, Vol. 22, No. 8, pp.3271–3282.

He, K., Sun, J. and Tang, X. (2009) 'Single image haze removal using dark channel prior', *Computer Vision and Pattern Recognition 2009 IEEE Conference*, Florida.

Krizhevsky, A., Sutskever, I. and Hinton., G.E. (2012) 'Imagenet classification with deep convolutional neural networks', *Advances in Neural Information Processing Systems*, Vol. 1, pp.1097–1105.

LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W. and Jackel, L.D. (1989) 'Backpropagation applied to handwritten zip code recognition', *Neural Computation*, Vol. 1, No. 4, pp.541–551.

Lee, S., Yun, S., Nam, J-H., Won, C.S. and Jung, S-W. (2015) 'A review on dark channel prior based image dehazing algorithms', *EURASIP Journal on Image and Video Processing*, Vol. 4, doi:10.1186/s13640-016-0104-y

Li, Y., Huang, J. and Luo, J. (2015) 'Using user generated online photos to estimate and and monitor air pollution in major cities', *International Conference on Internet Multimedia Computing and Service (ICIMCS 2015)*, doi: http://dx.doi.org/10.1145/ 2808492.2808564.

Liu, F., Shen, C. and Lin, G. (2015) 'Deep convolutional neural fields for depth estimation from a single image', *IEEE Conference on Computer Vision and Pattern Recognition*, Boston.

Padole, N. and Khare, A. (2015) 'Improved method of single image dehazing based on multi-scale fusion', *International Journal of Computer Science and Information Technologies*, Vol. 6, No. 3, pp.2945–2949.

Saggu, M.K. and Singh, S. (2015) 'A review on various haze removal techniques for image processing', *International Journal of Current Engineering and Technology*, Vol. 5, No. 3, pp.1500–15005.

Sreekuttan, A.M. (2016) *Single Image Dehazing Using Multiple Fusion Technique*, MatWorks [online] https://www.mathworks.com/matlabcentral/fileexchange/47147-image-dehazing-zip (accessed 6 February 2017).

Wang, Z., Bovik, A.C., Sheikh, H.R. and Simoncelli, E.P. (2004) 'Image quality assessment: from error measurement to structural similarity', *IEEE Transactions on Image Processing*, Vol. 13, pp.600–612.

Woodell, G., Jobson, D.J., Rahman, Z-U. and Hines, G. (2006) 'Advanced image processing of aerial imagery', *Visual Information Processing XV*, Vol. 6246, 62460E.

Zhengguo, D. (2012) *A New Visibility Metric For Haze Images*, MathWorks [online] https://www.mathworks.com/matlabcentral/fileexchange/33529-a-new-visibility-metric-for-haze-images (accessed 6 February 2017).