Ministry of Education of the Republic of Belarus
Educational institution
Belarusian State University of
Informatics and Radioelectronics

UDC 004.42:629.33

Nayyef Mohanad Nidhal Nayyef

Video Sequence Filtering Based On Convolutional Neural Network

ABSTRACT
For master’s degree in technical sciences

Specialty 1-45 80 02 «Telecommunication systems and computer networks»

____

Supervisor
Baryskievic I.A
PhD, associated professor

____

Minsk 2019

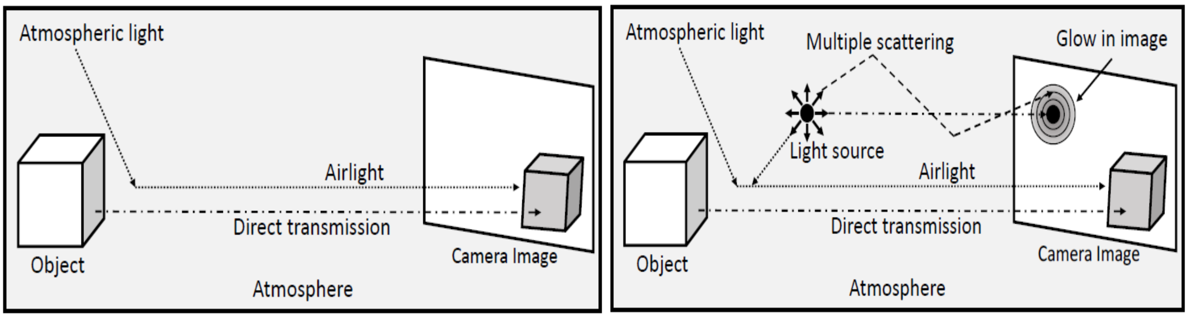
INTRODUCTION

The presence of haze significantly degrades the quality of an image captured at night. Similar to daytime haze, the appearance of nighttime haze is due to tiny particles floating in the air that adversely scatters the line of sight of lights rays entering the imaging sensor. In particular, light rays are scattered-out to directions other than the line of sight, while other light rays are scattered-in to the line of sight. The scattering-out process causes the scene reflection to be attenuated. The scattering-in process creates the appearance of a particles-veil (also known as air light) that washes out the visibility of the scene. These combined scattering effects adversely affect scene visibility that in turns negatively impacts subs equation sent processing for computer vision applications.

In contrast to these methods, we model nighttime haze images by explicitly taking into account the glow of active light sources and their light colors. This new model introduces a unique set of new problems, such as how to decompose the glow from the rest of the image and how to deal with varying atmospheric light. By resolving these problems, we found our results are visually more compelling than both existing daytime and nighttime methods.

Figure a, shows a diagram of the standard daytime haze model. The model assumes that the atmospheric light is globally uniform and contributes to the brightness of the air light. The model has another term called the direct transmission, which describes light travelling from the object or scene reflection making its way to the image plane.

Figure b, shows a diagram of our proposed nighttime haze model. Aside from the air light and direct transmission, the model also has a glow term, which represents light from sources that gets scattered multiple times and reaches the image plane from different directions. In our model, light sources potentially have different colors that contribute to the appearance of the air light.



a b

a – Daytime haze imaging model; b – Nighttime haze imaging model
diagram of the daytime and nighttime haze models

GENERAL DESCRIPTION OF WORK

The proposed algorithm focuses on dehazing nighttime images. Most existing dehazing methods use models that are formulated to describe haze anytime. Daytime models assume a single uniform light color attributed to a light source not directly visible in the scene. Nighttime scenes, however, commonly include visible lights sources with varying colors. These light sources also often introduce noticeable amounts of glow that is not present in daytime haze. To address these effects, we introduce a new nighttime haze model that accounts for the varying light sources and their glow. Our model is a linear combination of three terms: the direct transmission, air light and glow. The glow term represents light from the light sources that is scattered around before reaching the camera. Based on the model, we propose a framework that first reduces the effect of the glow in the image, resulting in a nighttime image that consists of direct transmission and air light only. We then compute a spatially varying atmospheric light map that encodes light colors locally. This atmospheric map is used to predict the transmission, which we use to obtain our nighttime scene reflection image. We demonstrate the effectiveness of our nighttime haze model and correction method on a number of examples and compare our results with existing daytime and nighttime dehazing methods' results.

Algorithm for single image haze removal using the dark channel prior.

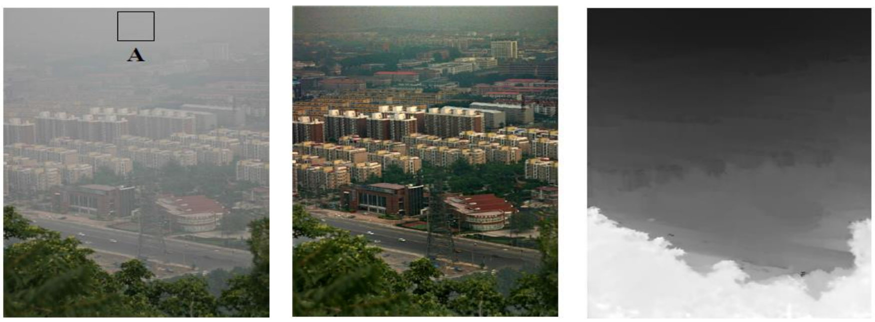
Haze Imaging Model. The haze imaging equation:

$$I(x) = J(x)t(x) + A(1 - t(x)).$$

where $x = (x, y)$ is a 2D vector representing the coordinates (x, y) of a pixel's position in the image.

I represent the hazy image observed. $I(x)$ is a 3D RGB vector of the color at a pixel.

J represents the scene radiance image. $J(x)$ is a 3D RGB vector of the color of the light reflected by the scene point at x. It would be the light seen by the observer if this light were not through the haze. So we often refer to the scene radiance J as a haze-free image



I J t

Figure 1 – Variables in the haze imaging equation. The transmission map t is shown as white when t=1, and black when t=0.

The neural network for removal of haze

The neural network for removal of HAZE. The atmospheric scattering model in Section II-A suggests that estimation of the medium transmission map is the most important step to recover a haze-free image. To this end, we propose DehazeNet, a trainable end-to-end system that explicitly learns the mapping relations between raw hazy images and their associated medium transmission maps. In this section, we present layer designs of DehazeNet, and discuss how these designs are related to ideas in existing image dehazing methods. The final pixel-wise operation to get a recovered haze-free image from the estimated medium transmission map will be presented in Section IV.

A.Layer Designs of DehazeNet. The proposed DehazeNet consists of cascaded convolutional and pooling layers, with appropriate nonlinear activation functions employed after some of these layers. shows the architecture of DehazeNet. Layers and nonlinear activations of DehazeNet are designed to implement four sEquation uentlyial operations for medium transmission estimation, namely, feature extraction, multi-scale mapping, local extremum, and nonlinear regression.

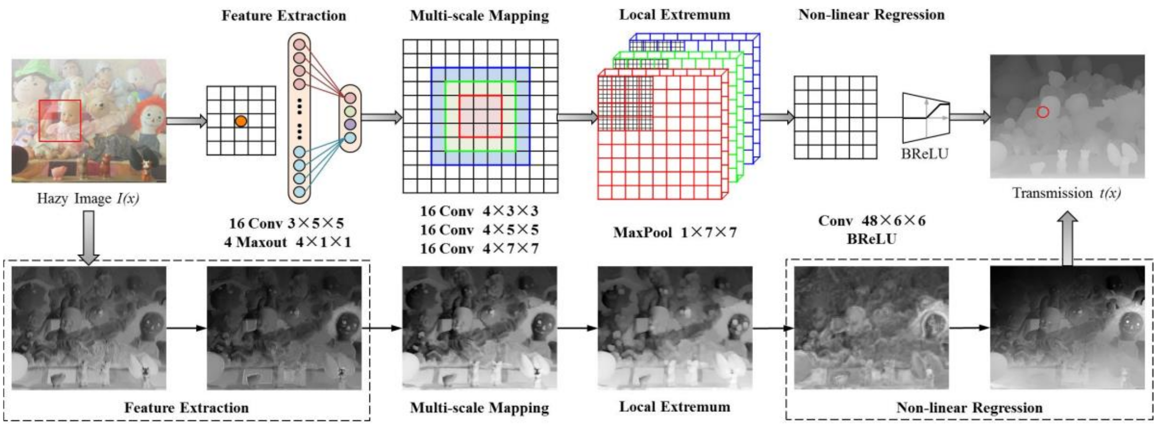


Figure 2 – The architecture of DehazeNet. DehazeNet conceptually consists of four equations Consequently, operations (feature extraction, multi-scale mapping, local extremum and non-linear regression), which is constructed by 3 convolution layers, a max-pooling, a Maxout unit and a BReLU activation function.

The algorithm for image dehazing based on MLP

The atmospheric scattering model. The atmospheric scattering is a physical phenomenon where the light passing through the particles in the atmosphere is deviated from its straight path. The formation of an image can be explained using the atmospheric scattering model proposed by McCartney et al., and it is defined as follows:

$$I(x,y) = J(x,y)t(x,y) + A(1 - t(x,y)),$$

where $I(x,y)$ corresponds to the captured image, $J(x,y)$ is the image of the original scene without affectations, A is the color of atmospheric light, and $t(x,y)$ is called transmission, which can be defined in a homogeneous atmosphere as:

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

where β is the scattering coefficient of the atmosphere and $d(x,y)$ is the scene depth.

In order to get a haze-free image, the equation can be expressed as follows:

$$J(x,y) = \frac{I(x,y) - A}{t(x,y)} + A$$

It is important to point out that in Equation [119], additionally to the input image $I(x,y)$, the transmission map $t(x,y)$ and atmospheric light A must to be determined . The DCP method makes possible to obtain an accurate estimation of the variables $t(x,y)$ and A.

The training step of a MLP requires quires two vectors of samples: the input data and the target vectors. In this work, each sample was acquired from a square window of size l centered in positions (x,y) where the length for each sample is $l \times l$. The input vector was obtained from $t_{\min}(x,y)$ and the target vector from $\bar{t}(x,y)$. The setup to perform the training stage is illustrated in Figure 4.3.

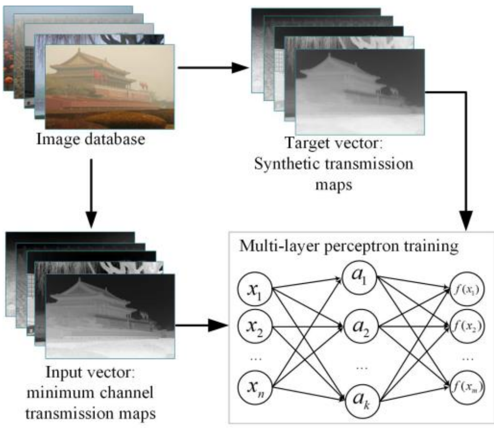
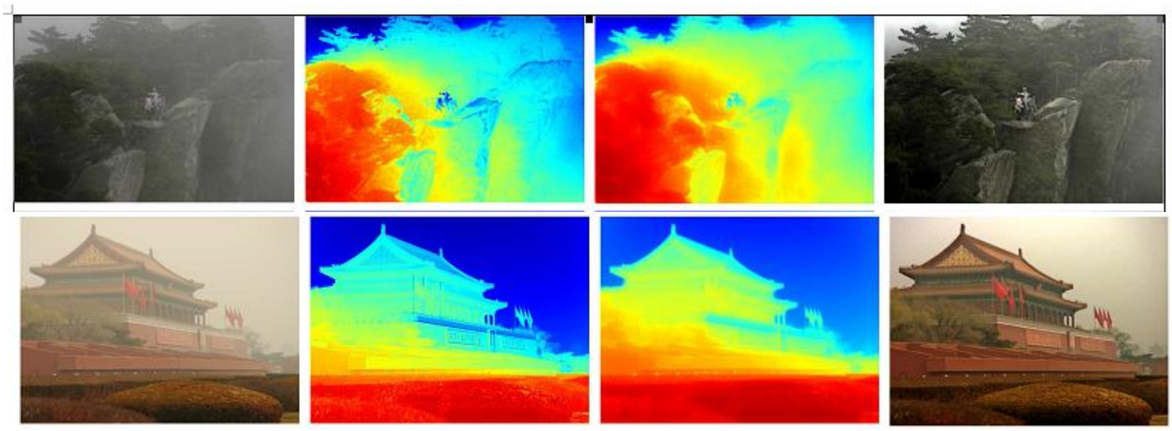


Figure 3 – Training process for the MLP based on transmission maps.

Application of MLP

Application of MLP. As it is shown in, to compute the transmission map $t'(x,y)$ using the trained MLP, an input vector is generated from square-windows of $t_{\min}(x,y)$. Each two-dimensional square-window of size s is converted to a one-dimensional signal of size $s \times s$. The interval of each sampling in the image is expressed by the variable δ. If $\delta = s$ the sampling has not intersection in the pixels positions, and if δ

= 1 the sampling is realized in every pixel. The value of each pixel in the transmission map $t'(x,y)$ is the minimum value of the superposition of windows.
In Figure 4.5 an example of the proposed algorithm in four real-world images and the results of applying the proposed method are presented, along with the input $t_{\min}(x,y)$ and output $t'(x,y)$ transmission maps of the MLP are shown. The parameters values are $\delta = 16$ and $s = 8$.



a b c d

Figure 4 – Examples of the processing the proposed method: (a) - input images $I(x,y)$, (b) - initial transmission $t_{\min}(x,y)$ (c) - final transmission $t'(x,y)$, (d) - recovered images $J(x,y)$.

Eval metrics. In order to measure the performance of the proposed method, two different metrics have been adopted: Mean Absolute Error (MAE), and Structural Similarity (SSIM) index.

The Mean Absolute Error (MAE) frequires Consequently,ly used in statistics provides a measure of disparity between the restored image $I^o(x,y)$ and the target image $J(x,y)$ using the width (ω) and height (h) of the input image, it is defined as:

$$MAE = \frac{1}{(\omega * h)} \sum_{x=1}^{\omega} \sum_{y=1}^{h} |J(x,y) - I^o(x,y)| , \tag{4.16}$$

The Structural similarity index (SSIM) is a normalized metric based on a perception-based model, and it is defined as:

$$S(x,y) = f(l(x,y), c(x,y), s(x,y)), \tag{4.17}$$

where $l(x,y)$ is the luminance comparison, $c(x,y)$ is the contrast comparison, and $s(x,y)$ is the structure comparison.

The eval performance of the proposed method was assessed using a database of 120 rgb images. The database was divided into two subsets: a training set of 80 images and a test set of 40 images. The test set of images consists of 30 real-world images and 10 synthetic images obtained from Fattal et al. [30] Furthermore, the computational experiments were performed using the Matlab software, version R2016a on a computer with 3.10 GHz Intel Core i5-2400 and 12 GB RAM memory.

Parameters tuning. To obtain the best possible performance of the proposed method, several tests with different MLP configuration s using the training set of 80 images were realized. In Figure 4.6, the results of the most significant configurations in terms of the MSE metric are shown. The best MLP performance is presented by the configuration $256 - 1024$, which corresponds to a window size $s = 16$.

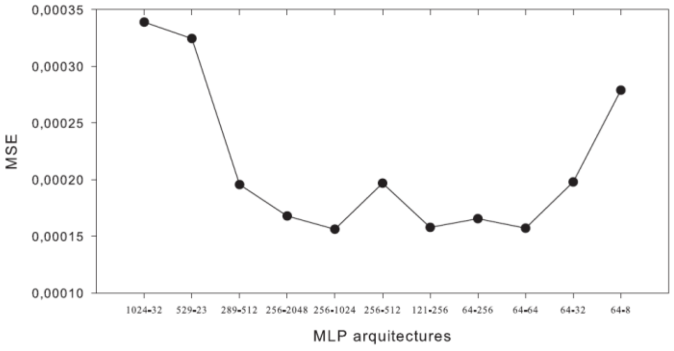
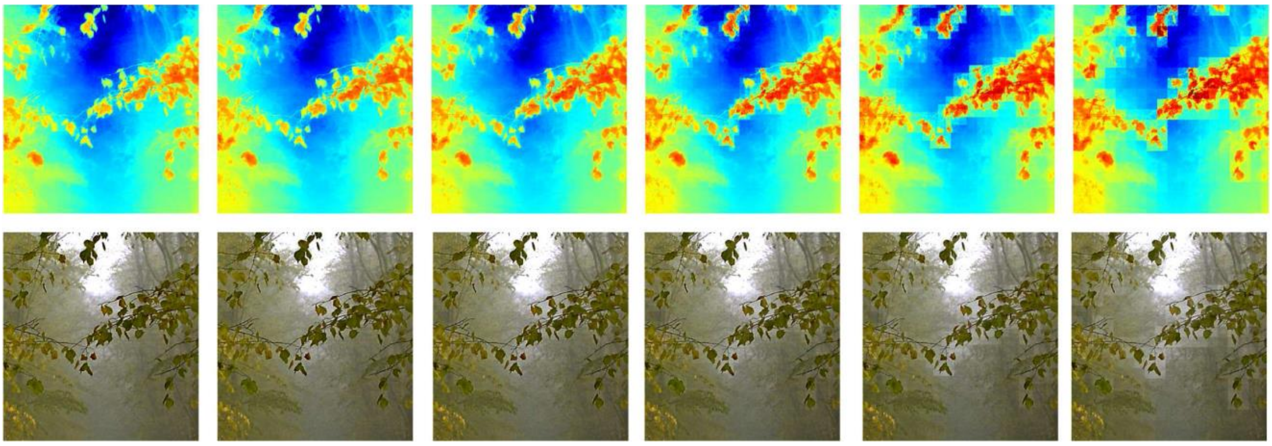


Figure 5 – Relation between MLP architecture and performance using the training set of images.

Parameters tuning

Once the value s was determined, in order to obtain the optimal value of δ (window sliding value), the method was tested over the training set varying the δ value in the range [1, .., 16]. Three aspects were considered to choose the δ value: a quantitative analysis using the index SSIM (Figure 4.7 (a)), a time processing analysis (Figure 4.7 (b)), and a qualitative analysis(Figure 4.8).Based on the obtained results, the best value was δ = 8, representing the highest SSIM and lowest processing time.



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

Figure 6 – Transmissions and results generated using different δsliding windows:
(a) δ = 1, (b) δ = 2, (c) δ = 4, (d) δ = 8, (e) δ = 12, (f) δ = 16.

CONCLUSION

We have presented a novel deep learning approach for single image dehazing. Inspired by traditional haze-relevant features and dehazing methods, we show that medium transmission estimation can be reformulated into a trainable end-to-end system with special design, where the feature extraction layer and the non-linear regression layer are distinguished from classical CNNs. In the first layer F_1, Maxout unit is proved similar to the priori methods, and it I more effective to learn haze-relevant features. In the last layer F_4, a novel activation function called BReLU is instead of ReLU or Sigmoid to keep bilateral restraint and local linearity for image restoration. With this lightweight architecture, DehazeNet achieves dramatically high efficiency and outstanding dehazing effects than the state-of-the-art methods. Although we successfully applied a CNN for haze removal, there are still some extensibility researches to be carried out. That is, the atmospheric light _ cannot be regarded as a global constant, which will be learned together with medium transmission in a unified network. Moreover, we think atmospheric scattering model can also be learned in a deeper neural network, in which an end-to-end mapping between haze and haze-free images can be optimized directly without the medium transmission estimation.

The single image dehazing method was proposed. This method uses an artificial neuronal network Multi-Layer Perceptron (MLP) to estimate the transmission map of a haze image. To obtain the optimal MLP conFigureuration a training set of 80 real-world images was used. In experiments a number of hidden layers containing different number of neurons was tested, where the best performance in terms of the Mean Squared Error (MSE) was achieved using a 256−1024 MLP conFigureuration with a MSE = 0.000151. In order to evaluate the restoration quality of the proposed method, the Structural Similarity (SSIM) index and the Mean Absolute Error (MAE) were used. The experimental results have proven that the proposed method achieves a superior performance than seven state-of-art methods in terms of restoration quality, obtaining aMAE value of 27.28 and SSIM index of 0.84 over a test set of synthetic images. In addition, a comparative analysis using the test set of 40 images between the proposed and comparative methods, reveals that the lowest computational time was obtained by the proposed method (0.52 seconds). Given the suitable results in terms of restoration quality and execution time of the proposed method with respect to the state-of-art dehazing methods, it can be highly appropriate to be used in real time systems.