creative commons
COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

이학박사 학위논문

Research On Moving Object Detection And Dusty Image Restoration

(이동 물체 감지 및 분진 영상 복원의 연구)

2021년 **2**월

서울대학교 대학원

수리과학부

안효민

Research On Moving Object Detection And Dusty Image Restoration

이동 물체 감지 및 분진 영상 복원의 연구

지도교수 강 명 주

이 논문을 이학박사 학위논문으로 제출함

2020 년 11 월

서울대학교 대학원

수리과학부

안 효 민

안효민의 이학박사 학위논문을 인준함

2020 년 11 월

위 원 장 국 웅 (인)

부위원장 강 명 주 (인)

위 원 이 기 암 (인)

위 원 정 미 연 (인)

위 원 강 명 민 (인)

Research On Moving Object Detection And Dusty Image Restoration

A dissertation
submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
to the faculty of the Graduate School of
Seoul National University

by

Hyomin Ahn

Dissertation Director : Professor Myungjoo Kang

Department of Mathematical Science
Seoul National University

February 2021

© 2020 Hyomin Ahn

All rights reserved.

Abstract

Robust principal component analysis(RPCA), a method used to decompose a matrix into the sum of a low-rank matrix and a sparse matrix, has been proven effective in modeling the static background of videos. However, because a dynamic background cannot be represented by a low-rank matrix, measures additional to the RPCA are required. In this thesis, we propose masked RPCA to process backgrounds containing moving textures. First-order Marcov random field (MRF) is used to generate a mask that roughly labels moving objects and backgrounds. To estimate the background, the rank minimization process is then applied with the mask multiplied. During the iteration, the background rank increases as the object mask expands, and the weight of the rank constraint term decreases, which increases the accuracy of the background. We compared the proposed method with state-of-art, end-to-end methods to demonstrate its advantages.

Subsequently, we suggest novel dedusting method based on dust-optimized transmission map and deep image prior. This method consists of estimating atmospheric light and transmission in that order, which is similar to dark channel prior-based dehazing methods. However, existing atmospheric light estimating methods widely used in dehazing schemes give an overly bright estimation, which results in unrealistically dark dedusting results. To address this problem, we propose a segmentation-based method that gives new estimation in atmospheric light. Dark channel prior based transmission map with new atmospheric light gives unnatural intensity ordering and zero value at low transmission regions. Therefore, the transmission map is refined by scattering model based transformation and dark channel adaptive non-local total variation (NLTV) regularization. Parameter optimizing steps with deep image prior(DIP) gives the final dedusting result.

Key words: Moving object detection, Dynamic background subtraction, Ro-

bust principal component analysis, Image dedusting, Deep image prior
Student Number: 2015-20267

Contents

Chapter 1

Introduction

1.1 Moving Object Detection In Dynamic Backgrounds

Background subtraction is a major approach to moving object detection and is required for automated video analysis and its practical applications in traffic monitoring, auto driving, and fire detection. The concept is to model a background without moving objects and compare it with each input video frame. Pixels with differences between them, which are greater than a set threshold, are considered moving objects.

It is of significance that the background is not always static. That is, there is a difference in the relative importance of individual moving objects, which means that sometimes only important objects should be extracted. This problem is termed dynamic background subtraction.

Robust principal component analysis (RPCA) [9] is a useful approach for static background subtraction owing to its ability to recover low-rank ma-

trices. That is, the low-rank output of the RPCA models the static background, and the sparse output models the moving object. However, as the low-rank output of the RPCA is literally close to the low-rank matrix, its ability to express the dynamic background is limited. To compensate for this limitation we introduced a mask to distinguish the object and background. We found that it supports modeling the dynamic background and selecting the major objects among the combination of objects. By combining the prevalent RPCA framework and the object mask, we produce another RPCA model for dynamic background subtraction.

The contributions of this thesis are as follows:

• We propose a new type of RPCA-based dynamic background subtraction method that differently emphasizes the similarities between frames according to the object mask.

• Our method does not require matrix decomposition such as singular value decomposition (SVD), and takes a near linear number of computations for each iteration with respect to the number of pixels in the video.

• Experiments show that our method outperforms state-of-art end-to-end non-deep learning-based methods.

We briefly present some related works in Section . The proposed method and its optimization approach are then introduced in Section 3.1 and 3.2. In Section 3.3, we present the detailed implementation of the algorithm and the result of the experiments demonstrating the superiority of our method.

1.2 Image Dedusting

Particulate matter, which is a kind of fine dust generated by factories, power plants, etc., can stay in the air for some days. Pictures taken in dusty environments have weak colors and low contrast. This is because the particles in the atmosphere scatter and absorb the light reflected from objects[39]. The

proportion of light not scattered and arriving at the camera is termed as transmission, and transmission depends on the distances from the objects in the picture to the camera. The degradation of images by dust hinders direct human use and the application of machine vision techniques. Therefore, a method to restore an image polluted by dust is required.
Fine dust is a recent phenomenon in East Asia, so the development of dedusting methods has been scarce. In this thesis, we propose a new image restoration method for dusty environments. The degradation model is based on the mechanisms of atmospheric scattering.

The contributions of this thesis are as follows:

• We compute atmospheric light with superpixel segmentation, which fits fine dust. Moreover, we refined the transmission map to fit the new atmospheric light.

• Based on deep image prior(DIP), we proposed a neural dedusting method, i.e., U-Net and fully connected networks are applied to capture the priors of clean image and unknown parameters, respectively.

• Experiments show that our method successfully recovers an image degraded by a dusty environment, and it outperforms other state-of-art dehazing methods.

The rest of the thesis is organized as follows. We present some related works in Section 2.2. Then introduce our proposed method in Section 4.1 and 4.2. In Section 4.3, we present the detailed implementation of the proposed algorithm and the results of the experiments compared with state-of-art dehazing algorithms. Finally, we conclude the thesis in Section 5.

Chapter 2

Preliminaries

In this section, the required preliminaries for this thesis is provided. We first introduce some related works for moving object detection in dynamic background. Subsequently, preliminaries for image dedusting is presented.

2.1 Moving Object Detection In Dynamic Backgrounds

In practice scenarios, the background contains moving textures that should be considered part of backgrounds. Usually, such textures are moving objects which are small or move slightly. Swaying branches, flowing water, and moving cars far from the camera are included in that category. Noise, camera jitter, change of lights, and pale shadows also contributes to the formation of such textures. Figure 2.1 shows examples of the dynamic background and moving object detection in it. The first low of Figure 2.1 shows the input frame. The first two columns are from "Escalator" and the letter two columns are from "WaterSurface". Both of them are from the I2R dataset

Figure 2.1: Moving object detection in dynamic backgrounds using the proposed method. The input frames (the first column) and background subtraction results (the second column).

[28]. "Escalator" contains moving escalator, and "WaterSurface" contains flowing river as the moving texture. However, they should be considered the part of the background, and the only moving objects should be detected is the people in both case. The second low of Figure 2.1 shows the background subtraction results with the proposed method.

2.1.1 Literature review

Statistical methods are steadily being developed, and one significant class of these methods models the pixels with Gaussian distributions. Pfinder[62] detects and tracks a human body using a single-Gaussian model. Stauffer and Grimson[53] proposed the Gaussian mixture model (GMM). In [53], the value of each pixel is modeled as a mixture of Gaussians. After the Gaussian distributions representing the background are determined, the pixels whose values are close to these distributions are considered the background. Zivkovic

proposed improved-mixture-of-Gaussians background model (IMOG)[71], which is an extension of the GMM. In IMOG, the parameters and the number of components of the mixture are adaptively updated using recursive equations.

Another popular class of statistical methods deals with texture features. Heikkila and Pietikainen[19] proposed a method to find background statistics using discriminative texture features. They chose a local binary pattern(LBP) texture operator[41] for background modeling. LBP is an eight bit binary number that describes the grayscale texture, and it is defined as

$$LBP(x_c) = \sum_{p=0}^{7} s(g_p - g_c)2^p, \tag{2.1}$$

where g_c is the gray value of the central pixel x_c and g_p is that of the neighboring pixels. Zhang *et al.* proposed STLBP[67], which extends LBP from the spatial domain to spatio-temporal domain. He *et al.* [18] considered spatio-temporal local compact binary patterns(STLCBPs) to detect appearance changes in complex scenes.

ViBe[1] is a seminal work that uses a background sample-based model. In this work, the value of each pixel in the current frame is compared with background samples, which are constructed by previous frames. Pixels whose number of matches is greater than the threshold are considered to be background. These background samples are updated using a random sample replacement policy. Hofmann *et al.* [20] presented dynamic controllers for the decision threshold and learning parameters. St-Charles *et al.* improved those controllers in SuBSENSE[52]. Mandal proposed CANDID[35], which employs a deterministic background sample update policy.

Sparse representation with RPCA is another popular method for background subtraction. RPCA decomposes a matrix into sum of a sparse matrix and a low-rank matrix. The underlying concept is that backgrounds have a lower rank than the original video with its moving objects, and the mov-

ing objects represent a small percentage of the total number of pixels in the video. However, as mentioned before, dynamic backgrounds can have high rank and additional measures are needed to process them.

The use of MRF can result in boundary adherence and sparsity, and it has been proved effective in enhancing the detection accuracy. Zhou *et al.* [69] and [22] modeled the object in front using first-order MRFs with binary labels. Javed *et al.* [23] used MRF for post-processing. Wang *et al.*[57] proposed a two-layer MRF model: one deals with a pixel-based constraint and the other with a superpixel-based constraint.

Recently, deep learning approaches such as [31] have demonstrated superior performance. However, they require massive training data, which are hard to acquire. In addition, if the style of the testing data is different from the training data, they do not demonstrate adequate performance, and in this case, it is difficult to tune their parameters without additional data.

2.1.2 Robust principal component analysis(RPCA) and their application status

Let $X \in \mathbb{R}^{d \times n}$ be a matrix. In the case of background subtraction, each column of this matrix is flattened frame of the input video. Then, the fundamental concept of the RPCA is to decompose X as

$$X = L + S, \tag{2.2}$$

where L is a low-rank matrix representing the background and S is a sparse matrix representing the moving objects. Because minimizing directly

$$\text{rank}(L) + ||S||_0, \text{ s.t., } L + S = X, \tag{2.3}$$

is known to be NP-hard, the condition for both the L and the S term should be relaxed. Candès *et al.* [9] suggested the most basic form of the RPCA method for this field by minimizing

$$||L||_* + \lambda||S||_1, \text{ s.t., } L + S = X, \tag{2.4}$$

where $||\cdot||_*$ is the nuclear norm. Bouwmans and Zahzah [4] reviewed algorithms with the ranks minimized by using nuclear norm minimization. Gu *et al.* [15] used the weighted nuclear norm minimization (WNNM) framework instead of nuclear norm minimization. Oh *et al.* [40] introduced partial sum minimization of singular values (PSSV), which is the sum of singular values except for the largest ones. A significant drawback of such methods is that they all require SVDs, of which the per-iteration complexity is $O(nmd)$ where $m = \min(d, n)$. Because $d >> n$ usually, the above algorithms become extremely slow for long videos. Therefore, many attempts have been made to reduce the time consumed by amending the rank term of L. LMaFit [50], Unifying [7], and factEN [24] suggest partial SVD schemes, of which the per-iteration complexity is $O(cnd)$ where $c << \min(d, n)$. They share the approach of decomposing the low rank component into $L = UV^T$, where $U \in \mathbb{R}^{n\times c}$, $V \in \mathbb{R}^{d\times c}$. LMaFit minimizes the l_1-norm of the difference between X and L. Besides this, Unifying minimizes the Frobenius norm of U and V. In addition, factEN further minimizes the Frobenius norm of L.

Shang *et al.* [49] noticed that the nuclear norm is identical to the Schatten 1-norm and generalized (2.4) by

$$||L||_{S^p}^p + ||S||_p^p, \text{ s.t., } L + S = X, \tag{2.5}$$

where $||\cdot||_{S^p}$ is the Schatten p-norm. Their study showed that, for $p = 1/2$ and $p = 2/3$, the Schatten p-norm can be transformed into a simple form that can be numerically minimized that outperformed those in [50],

[7], and [24]. Peng *et al.*[42] proposed RES-PCA which introduces another constraint for the rank term, which is entirely separate the singular values of X. Beyond the prevalent low-rank assumption of L, they assumed that the columns of L are highly approximate to each other in the Euclidean space. This led to the replacement of the rank constraint term with

$$\sum_{i=1}^{n}\sum_{j=1}^{n}||L_i - L_j||_{\mathrm{F}}^2 \tag{2.6}$$

or equally,

$$Tr\left(L(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^{\mathrm{T}})L^{\mathrm{T}}\right), \tag{2.7}$$

where $\mathbf{1}_n$ is an n-dimensional column vector containing 1's.

Now we briefly introduce the solver of each applications. The main contents of [9] is minimizing (2.4), and it is realized by augmented Lagrange multiplier(ALM) algorithm introduced in [32, 65]. The ALM method operates on the augmented Lagrangian

$$l(L,S,Y) = ||L||_* + \lambda||S||_1 + < Y, X - L - S > + \frac{\mu}{2}||X - L - S||_{\mathrm{F}}^2. \tag{2.8}$$

Its minimizing strategy is to first solve L-subproblem (fixing S), then solve S-subproblem (with L fixed), and then update the Lagrange multiplier matrix Y based on the residual $X-L-S$. Let $\mathcal{S}_\tau$ denote the shrinkage operator defined by $\mathcal{S}_\tau = \mathrm{sgn}(x)\max(|x| - \tau, 0)$. Then the S-subproblem is solved by

$$\underset{S}{\mathrm{argmin}}\ l(L,S,Y) = \mathcal{S}_{\lambda\mu}(X - L + \mu^{-1}Y). \tag{2.9}$$

On the other hand, for a matrix Z, let $\mathcal{D}_\tau(Z)$ denote the singular value

thresholding operator defined by $\mathcal{D}_\tau(Z) = U\mathcal{S}_\tau(\Sigma)V^*$, where $Z = U\Sigma V^*$ is any singular value decomposition. Then the L-subproblem is solved by

$$\underset{L}{\operatorname{argmin}}\ l(L, S, Y) = \mathcal{D}_\mu(X - S - \mu^{-1}Y). \tag{2.10}$$

Oh *et al.* [40] can be summarized in minimizing

$$\min_{L,S} ||L||_{p=N} + \lambda||S||_1, \text{ s.t. } X = L + S, \tag{2.11}$$

where $||L||_p = \sum_{i=p+1}^{\min(n,d)} \sigma_i(L)$ and $\sigma_i(L)$ is the i-th singular value of L. It is also solved by ALM algorithm, and its augmented Lagrangian function is formulated by

$$l(L, S, Y) = ||L||_{p=N} + \lambda||S||_1 + < Y, X - L - S > + \frac{\mu}{2}||X - L - S||_{\mathrm{F}}^2. \tag{2.12}$$

[40] proved that the solution of minimization problem

$$\underset{X}{\operatorname{argmin}} \frac{1}{2}||X - Y||_{\mathrm{F}}^2 + \tau||X||_{p=N} \tag{2.13}$$

can be expressed by the PSVT operator defined as:

$$\mathcal{P}_{N,\tau}[Y] = U_Y(D_{Y_1} + \mathcal{S}_\tau[D_{Y_2}])V_Y^{\mathrm{T}}, \tag{2.14}$$

where

$$\begin{aligned} D_{Y_1} &= \operatorname{diag}(\sigma_1, ..., \sigma_N, 0, ..., 0), \\ D_{Y_2} &= \operatorname{diag}(0, ..., 0, \sigma_{N+1}, ..., \sigma_{\min(n,d)}), \end{aligned} \tag{2.15}$$

and $U_Y(D_{Y_1} + D_{Y_2})V_Y^{\mathrm{T}}$ is a SVD of Y. Therefore, (2.12) can be minimized by updating S and L alternatively at each iteration i as:

$$\begin{aligned} L_{i+1} &= \mathcal{P}_{N,\mu_i^{-1}}[X - S_i + \mu_i^{-1}Y_i], \\ S_{i+1} &= \mathcal{S}_{\lambda\mu_i^{-1}}[X - L_{i+1} + \mu_i^{-1}Y_i]. \end{aligned} \tag{2.16}$$

We omit the update of Y and μ.

LMaFit [50] replaced the rank constraint $\mathrm{rank}(L) \leq k$ by $L = UV$, where $U \in \mathbb{R}^{n\times c}$ and $U \in \mathbb{R}^{c\times d}$. Their model becomes:

$$\min_{U,V,L} ||X - L||_1 \text{ s.t. } UV - L = 0, \tag{2.17}$$

where $U \in \mathbb{R}^{n\times c}$ and $U \in \mathbb{R}^{c\times d}$ for some given but adjustable rank estimate $c > 0$. This is also solved by minimizing augmented Lagrangian function which is defined as

$$l(U, V, L, Y) = ||X - L||_1 + < Y, UV - L > + \frac{\mu}{2}||UV - L||_{\mathrm{F}}^2. \tag{2.18}$$

U-subproblem and V-subproblem of (2.18) are simple least-square problems, whose solutions at i-th iteration are

$$B = L_i - \frac{L_i}{\mu}, \ U_{i+1} = BV_i^{\dagger}, \ V_{i+1} = U_{i+1}^{\dagger}B, \tag{2.19}$$

where $Z^{\dagger}$ denotes the pseudo-inverse of the matrix Z. The solution of L-subproblem can also be explicitly using the shrinkage operator, leading to

$$L_{i+1} = \mathcal{S}_{\mu^{-1}}(U_{i+1}V_{i+1} - X + \frac{Y}{\mu}) + X. \tag{2.20}$$

Once again, we omit the update process of Y.

Shang *et al.* [49] generalized the regularization of RPCA with the Schatten p-norm. They studied for the case of $p = \frac{1}{2}$ and $\frac{2}{3}$, but we state only the case of $p = \frac{1}{2}$ here. Their energy to be minimized is

$$||L||_{S^{1/2}}^{1/2} + ||S||_{1/2}^{1/2}, \text{ s.t., } L + S = X. \tag{2.21}$$

First, they defined double nuclear norm penalty $||Z||_{D-N}$ with of a matrix Z whose rank is at most $r \leq c$ with two factor matix $U \in \mathbb{R}^{n \times c}$ and $V \in \mathbb{R}^{d \times c}$ such that $Z = UV^{\mathrm{T}}$. The double nuclear norm penalty of Z is defined as

$$||Z||_{D-N} = \min_{U,V:Z=UV^{\mathrm{T}}} \frac{1}{4}(||U||_* + ||V||_*)^2. \tag{2.22}$$

They proved that the Schatten-1/2 quasi-norm is identical to the double nuclear norm penalty, i.e.,

$$||Z||_{S_{1/2}} = ||Z||_{D-N} = \min_{U,V:Z=UV^{\mathrm{T}}} \frac{1}{4}(||U||_* + ||V||_*)^2. \tag{2.23}$$

Replacing the Schatten-1/2 quasi-norm of (2.21) with double nuclear norm penalty, the problem is reformulated as

$$\begin{aligned} \min_{U,V,L,S,\hat{U},\hat{V}} & \frac{\lambda}{2}\left(||\hat{U}||_* + ||\hat{V}||_*\right) + ||S||_{l_{1/2}}^{1/2}, \\ & \text{s.t., } \hat{U} = U, \hat{V} = V, UV^{\mathrm{T}} = L, L + S = X. \end{aligned} \tag{2.24}$$

The alternating direction method of multipliers (ADMM, or the inexact ALM)

is used to solve (2.24), whose augmented Lagrangian function is given by

$$\begin{aligned} l_{mu}(U,V,L,S,\hat{U},\hat{V},Y_i) = \frac{\lambda}{2}(||\hat{U}||_* + ||\hat{V}||_*) + ||S||_{l_{1/2}}^{1/2} + < Y_1, \hat{U} - U > \\ + < Y_2, \hat{V} - V > + < Y_3, UV^{\mathrm{T}} - L > + < Y_4, L + S - X > \\ + \frac{\mu}{2}\left(||\hat{U} - U||_F^2 + ||\hat{V} - V||_{\mathrm{F}}^2 + ||UV^{\mathrm{T}} - L||_{\mathrm{F}}^2 + ||L + S - X||_{\mathrm{F}}^2\right). \end{aligned} \tag{2.25}$$

In this case, the $\hat{U}$-subproblem and the $\hat{V}$ subproblem are the sum of the nuclear norm and the Frobenius norm, therefore they can be solved by the singular value thresholding operator $\mathcal{D}_\tau$. Both the U-subproblem and the V-subproblem are least square problems, and their optimal solutions are given by

$$\begin{aligned} U_{i+1} &= (\hat{U}_i + \mu_i^{-1}Y_1^i + M_iV_i)(I_c + V_i^{\mathrm{T}}V_i)^{-1}, \\ V_{i+1} &= (\hat{V}_i + \mu_i^{-1}Y_2^i + M_i^{\mathrm{T}}U_{i+1})(I_c + U_{i+1}^{\mathrm{T}}U_{i+1})^{-1}, \end{aligned} \tag{2.26}$$

where $M_i = L_i - \mu^{-1}Y_3^i$, and I_c denotes the identity matrix of size c.

RES-PCA [42] directly imposed the similarity between columns in Euclidean space. Replacing the rank term with (2.7), they made the columns be included in a rank-1 subspace. For the case of L of rank c, they proposed multiple rank-1 subspaces with the following model:

$$\begin{aligned} \min_{L,S,p_1,\ldots,p_c} \lambda \sum_{i=1}^{c} Tr\left(L\mathbf{d}(p_i)(I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^{\mathrm{T}})\mathbf{d}(p_i)L^{\mathrm{T}}\right) + ||S||_1 \\ \text{s.t., } X = L + S,\ p_i \in 0,1^n,\ \sum_i p_i = \mathbf{1}_n. \end{aligned} \tag{2.27}$$

This can be solved by ALM algorithm.

Note that all of them are solved by ALM algorithms, and [9] and [40] re-

quires SVDs whose computational complexity is $O(nmd)$ where $m = \min(d, n)$. Although the per-iteration complexity of [50],[49], and [42] is $O(cnd)$ where $c > 0$ is the upper bound of rank of L, their L's are undesirable to model dynamic backgrounds since they are literally low-ranked.

2.1.3 Graph cuts and α-expansion algorithm

In this section, Bayesian labeling of first-order Markov Random Fields and graph cut algorithms to solve them are introduced. For detailed explanations, we refer readers to [6].

Suppose every pixel $p \in X$ must be assigned a label in some finite set $\mathcal{L}$. These labels tend to be piecewise smooth so they vary smoothly on the surface of an object, but can change dramatically at boundaries. The problem is to find a labeling f that assigns each pixel $p \in X$ a label $f_p \in \mathcal{L}$ while minimizing the energy whose form is

$$E(f) = E_{smooth}(f) + E_{data}(f), \tag{2.28}$$

where E_{smooth} measures how much the labeling f is not smooth, and E_{data} measures the disagreement between f and the given data.

Usually, $E(f)$ takes the form of

$$E(f) = \sum_{\{p,q\} \in \mathcal{N}} V_{p,q}(f_p, f_q) + \sum_{p \in X} D_p(f_p), \tag{2.29}$$

where $\mathcal{N}$ is the set of adjacent pairs of pixels, $V_{p,q}$ measure the distance between f_p and f_q, and D_p measures how well f_p does not fit pixel p given the observed data. $V_{p,q}$ is assumed to be a metric and D_p is non-negative but otherwise arbitrary.

For example, see Figure 2.2. In Figure 2.2.(a), a binary image is given.

Then a strong Gaussian noise is added to (a), and this is (b). Let x_p be the value of (b) at index p. (b) can be restored by minimizing the energy

$$E(f) = \sum_p D_p(f_p) + \sum_{(p,q)\in\mathcal{N}} 50|f_p - f_q|. \tag{2.30}$$

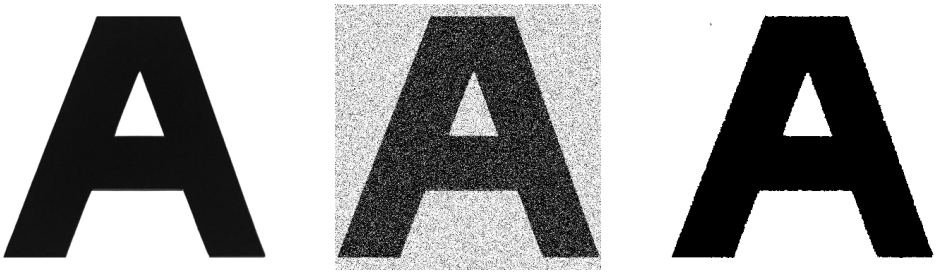
where $f_p \in \{0, 1\}$ for every p. Here, D is defined by

$$D_p(f_p) = \begin{cases} x_p & \text{if } f_p = 0 \\ 255 - x_p & \text{otherwise.} \end{cases} \tag{2.31}$$

After minimizing (2.30) with α-expansion algorithm explained below, one can get (c) for Figure 2.2.

A labeling f can be represented by a partition $\mathbf{P} = \{\mathcal{P}_l | l \in \mathcal{L}\}$, where $\mathcal{P}_l = \{p \in X | f_p = l\}$. Given a label α, a move from a partition $\mathbf{P}$ to $\mathbf{P}'$ is called an α-expansion if $\mathcal{P}_\alpha \subseteq \mathcal{P}'_\alpha$ and $\mathcal{P}'_\beta \subseteq \mathcal{P}_\beta$ for every $\beta \in \mathcal{L} - \{\alpha\}$. Then the α-expansion algorithm for minimizing (2.29) is shown in algorithm 1.

In the above algorithm, the method to find the optimal expansion move in step 3.a should be explained. [6] proved that the lowest energy labeling



(a) Ground truth (b) Corrupted image (c) Restored image

Figure 2.2: (a) Ground truth clean image, (b) (a) corrupted with additive Gaussian noise, (c) (b) restored by minimizing (2.30).

Algorithm 1 α-expansion algorithm

1: Start with an arbitrary labeling f
2: Set success $\leftarrow 0$
3: For each label $\alpha \in \mathcal{L}$
   a: Find $\hat{f} = \arg\min E(f')$ among f' within one α-expansion of f
   b: If $E(\hat{f}) < E(f)$, $f \leftarrow \hat{f}$ and success $\leftarrow 1$
4: If success $= 1$, go back to 2
5: Return f

within a single α-expansion move from f is $f^{\mathcal{C}}$, where $\mathcal{C}$ is the minimum cut on $\mathcal{G}_\alpha$. Then, this can be solved by $s-t$ cut algorithms which terminate in $\tilde{O}(n)$ times; [43] for example.

2.2 Image Dedusting

2.2.1 Image dehazing methods

The light scattering and absorbing phenomenon occur due to atmospheric particles. There have been many studies related to aerosols whose radius ranges from 10^{-2} to 1 μm. [10, 68] used multiple images to obtain diverse atmospheric conditions. This is not practical in real-world applications because it is hard to expect subjects not to change until the atmospheric environment changes. Therefore, the need for single image dehazing algorithms is raised.

The majority of single image dehazing algorithms have used the haze model [25]

$$I(x) = J(x)t(x) + A(1 - t(x)), \tag{2.32}$$

where I is the input hazy image, J is the latent scene radiance, A is the atmospheric light, t is the transmission, and x is the position of the pixel in the image. Transmission $t(x)$ can be formulated as:

$$t(x) = e^{-\beta D(x)}, \tag{2.33}$$

where $D(x)$ is the distance between the scene at the position and β is the scattering coefficient of the atmosphere. Because finding J from (2.32) is a highly ill-posed problem, there have been many attempts to solve (2.32) by utilizing a variety of prior and empirical knowledge. One remarkable approach among them is dark channel prior [17] based approach. The dark channel prior is based on observation on outdoor images. It says that in most non-sky patches of a clean outdoor image J, at least one color channel has some pixels whose intensity is very low. That is, for each non-sky patch $\Omega(x)$ centered at x,

$$DC_J(x) := \min_{y \in \Omega(x) \cap J} \left(\min_{c \in \{R,G,B\}} J^c(y) \right) \to 0. \tag{2.34}$$

where DC_J is called the dark channel of J. Dark channel prior based dehazing algorithms have been developed in the studies [14, 21, 47, 54, 70]. Figure 2.3 shows outdoor images(upper row) and their dark channel(lower row). Note that the rightmost picture is taken under a hazy environment and its dark channel value is large at hazy regions even though they not parts of the sky.

Another notable approach is learning-based methods. Dehazenet [8] used convolutional neural networks(CNN) framework to estimate transmission. Aodnet [27] directly generates a clean image through a light-weight CNN. Recently, the concept of generative adversarial networks(GAN) has been used by state-of-art methods, including GFN [44], Dehaze GAN [29], DCPDN

[66], DHSGAN [34], and Cycle-Dehaze [11].

2.2.2 Dust model

When the cause of scattering and absorbing is dust whose particle size is usually greater than 10μm, its scattering pattern is different from that of haze [38], and studies for restoring such images are very scarce. A dust imaging model considering its scattering pattern was suggested in [30, 60]:

$$I(x) = J(x)t(x)\left[1 + \frac{p(0,0)}{4\pi}\tau(x)\right] + A(1 - t(x)), \tag{2.35}$$

where p is the scattering phase function for dust, $\tau(x) = \beta D(x)$, and $D(x)$ is the distance between the scene at position x and the camera. When the

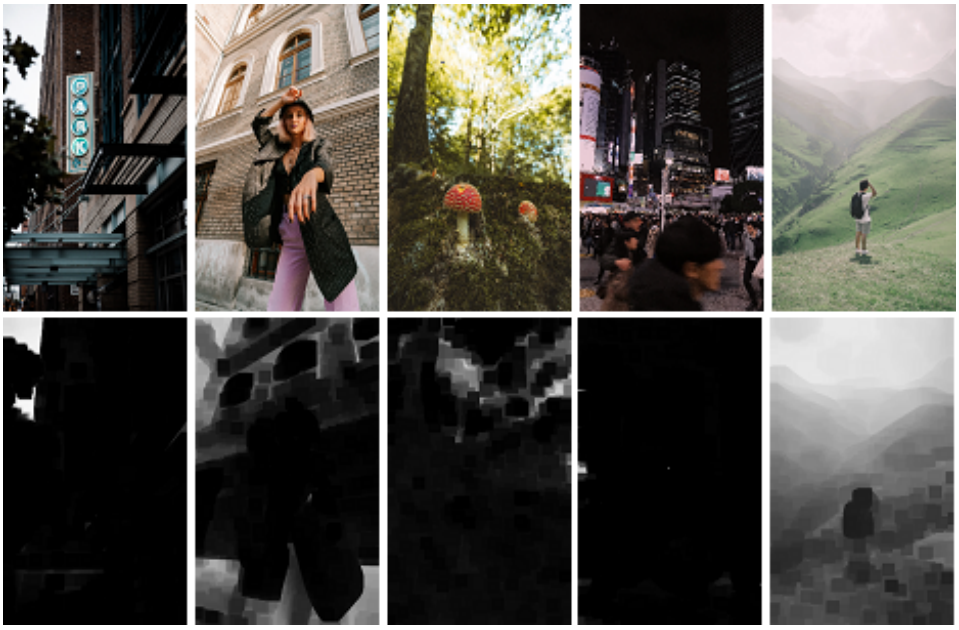


Figure 2.3: Images and their dark channels.

dust is homogeneous, transmission can be expressed as

$$t(x) = \exp\{-\beta D(x)\}. \tag{2.36}$$

Therefore, the dust model becomes

$$I(x) = J(x)t(x)\left[1 - \frac{p(0,0)}{4\pi}\log t(x)\right] + A(1 - t(x)). \tag{2.37}$$

2.2.3 Non-local total variation(NLTV)

The non-local similarity benefits analyzing complicated image structures. [13] introduced NLTV and showed minimizing NLTV can restore global regularity of images. Assume that a image t is given. In NLTV, a fixed area $A(x)$ is introduced for each pixel x. For each $y \in A(x)$, searching window $P(y)$ is a patch centered y. Then similarity between $P(x)$ and $P(y)$ is defined by

$$w(x,y) = \exp\left(\frac{1}{\sigma^2}||P(x) - P(y)||_G^2\right), \tag{2.38}$$

where G is a 2 dimensional Gaussian kernel matrix with standard deviation σ. Then the non-local gradient of t at x is define by

$$\nabla_w t(x) = (t(y) - t(x))\sqrt{w(x,y)}, \ \ \forall y \in A(x). \tag{2.39}$$

The NLTV is defined by

$$\sum_x |\nabla_w t(x)| = \sum_x \sqrt{\sum_{y \in A(x)} (t(y) - t(x))^2 w(x,y)}. \tag{2.40}$$

We applied NLTV minimization to refine the transmission map.

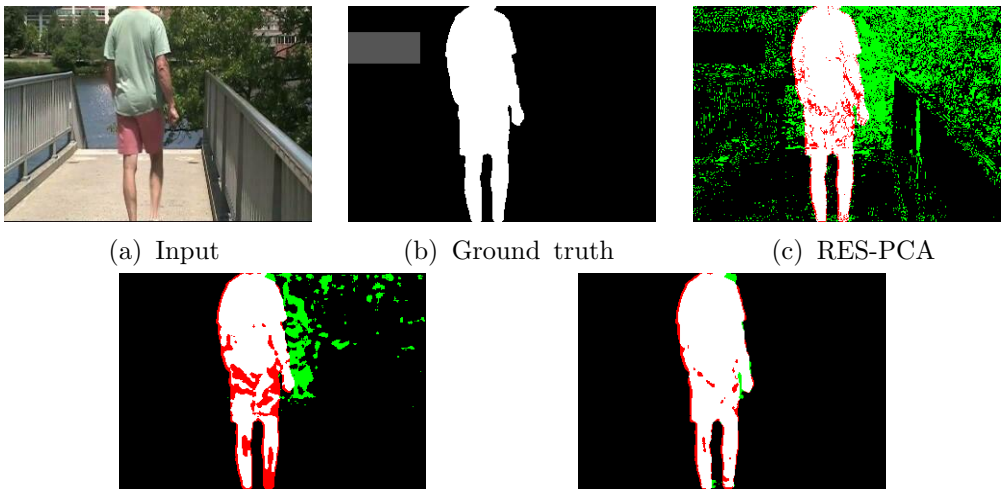
Chapter 3

Dynamic Background Subtraction With Masked RPCA

3.1 Motivation

3.1.1 Motivation of background modeling

Our assumption concerning the background was profoundly inspired by the work of RES-PCA[42], but [42] has two main points to improve. The first is the suboptimal rank of the background that makes it hard to represent dynamic backgrounds. To solve this problem, we separated L and $\tilde{L}$ where the former represents a dynamic background and the latter a static background. The second point that requires improvement is that minimizing (2.7) can cause ghost artifacts. According to [51], a ghost artifact is "a set of connected points, detected as being in motion but not corresponding to any real moving object." They may occur at pixel k if the proportion of j is high where X_{kj} is part of an object. To improve this, we formulated a bi-



(a) Input (b) Ground truth (c) RES-PCA

(d) minimizer of (3.6) without multiplication (e) Proposed method

Figure 3.1: Comparison with RES-PCA and the effect of mask multiplication. (a) Input frame, (b) ground truth (the gray region is the out-of-interest region), and background subtraction results of (c) RES-PCA, (d) minimizer of (3.6) without multiplying $(1 - M_j + \epsilon)$ at the first summation, and (e) the proposed method. (c) produced many false positives(green region) because RES-PCA gave excessively low-rank background and (d) produced ghost effects at the right side of the subject.

nary mask that represents moving objects. It was multiplied to the rank term for the background to be constructed from the regions without moving objects. A comparison with RES-PCA and the effect of mask multiplication is shown in Fig 3.1.

(2.6) can be transformed into

$$\sum_{i=1}^{n} ||L_i - \frac{1}{n}\sum_{j=1}^{n} L_j||_{\mathrm{F}}^2, \tag{3.1}$$

and this can be interpreted to mean that there is a vector $\tilde{L}$ where each

L_i is similar to it in Euclidean space. To apply the object mask yet to be introduced in section 3.1.2, we split $\tilde{L}$ from L. Therefore we changed (3.1) to

$$\min_{L,\tilde{L}} \sum_{i=1}^{n} ||L_i - \tilde{L}||_{\mathrm{F}}^2. \tag{3.2}$$

To cope with more general situations that may occur, such as sudden changes in lighting, we split the column space of L into c rank 1 subspaces and minimize the energy term given above for each subspace. That approach, when applied gives

$$\min_{L,p_1,\tilde{L_1},\ldots,p_c,\tilde{L_c}} \sum_{i=1}^{c} \sum_{(p_i)_j=1} ||L_j - \tilde{L_i}||_{\mathrm{F}}^2. \tag{3.3}$$

In this case, $\hat{L}$ where $\hat{L_j} = \tilde{L_i}$ for each j and i satisfies $(p_i)_j = 1$ is a low-rank matrix that represents the static background. We want L to represent the dynamic background, but the above formulation alone is insufficient to achieve this purpose. The background portion of each L_i must be rendered identically to that of X_i. To do this, we introduce a mask to represent objects in Section 3.1.2.

3.1.2 Mask formulation

Let $M \in \{0,1\}^{d \times n}$ be a matrix for distinguishing the foreground and the background, where

$$M(i,j) = \begin{cases} 1, & \text{if } ij \text{ is a moving object pixel} \\ 0, & \text{otherwise.} \end{cases} \tag{3.4}$$

Then the sparsity condition on S can be implemented by imposing sparsity on M. Now, let the object mask be the region where $M = 1$ and share its notation with M. Moreover, let the background mask be the region $M = 0$ and denote it as M^c. Assuming that the foreground objects are clustered and L and X are identical at M^c, we formulate the energy for M using the first-order MRF:

$$\sum_{j=1}^{n} \left(\beta ||M_j||_1 + \gamma \sum_{(kj,lj)\in\mathcal{N}} |M_{kj} - M_{lj}| \right) + \frac{\rho}{2} ||(1 - M) \circ (L - X)||_{\mathrm{F}}^2. \tag{3.5}$$

where $(kj, lj) \in \mathcal{N}$ indicates that pixels k and l are separated by a distance of 1 in the coordinates when the j-th column is converted to the original j-th video frame, and $\circ$ means component wise multiplication. We used $(kj, lj) \in \mathcal{N}$ instead of $(p, q) \in \mathcal{N}$; we focus only on the spatial smoothness of M. Otherwise the graph will be extremely large and the problem becomes difficult to solve. Note that the latter term of (3.5) can split into the sum of the energy of the each frame.

3.1.3 Model

$\tilde{L}_i$ obtained by optimizing (3.3) is the average of corresponding L_js, which is effected by moving objects. If ghost artifacts occur, they affect L and M sequentially, causing the algorithm to converge on poor results. Therefore, background modeling should be done by focusing on information from regions of M^c rather than dealing with information found in all pixels equally.

Combining this idea and (3.3) and (3.5), we propose the following model:

$$\begin{aligned}
&\min_{L,\tilde{L},M,p} \frac{\lambda}{2}\sum_{i=1}^{c}\sum_{(p_i)_j=1} ||(1-M_j+\epsilon)\circ(L_j-\tilde{L}_i)||_{\mathrm{F}}^2 \\
&+\frac{\rho}{2}||(1-M)\circ(L-X)||_{\mathrm{F}}^2 \\
&+\beta||M||_1+\gamma\sum_{j=1}^{n}\sum_{(kj,lj)\in\epsilon}|M_{kj}-M_{lj}| \\
&\quad s.t.\ p_i\in\{0,1\}^n,\ \sum_i p_i=\mathbf{1}_n,
\end{aligned} \tag{3.6}$$

where $\mathbf{d}(\cdot)$ is a diagonal matrix corresponding to the input vector, and $\mathbf{1}_n$ is an $n\times 1$ column vector with all components of 1. This energy is not convex and includes discrete variables, and, therefore, joint optimization over $L, \tilde{L}, M,$ and p is difficult. Consequently, we use an algorithm that alternatively solves the subproblem of each variable.

3.2 Optimization

3.2.1 L-Subproblem

L-subproblem is given by

$$\begin{aligned}
\min_{L}\frac{\lambda}{2}\sum_{i=1}^{c}\sum_{(p_i)_j=1}||(1-M_j+\epsilon)\circ(L_j-\tilde{L}_i)||_{\mathrm{F}}^2 \\
+\frac{\rho}{2}||(1-M)\circ(L-X)||_{\mathrm{F}}^2.
\end{aligned} \tag{3.7}$$

It is easily solved for each j and i with $(p_i)_j = 1$ as follows:

$$L_j = (1 - M_j) \circ \frac{(1+\epsilon)\lambda\tilde{L}_i + \rho X_j}{(1+\epsilon)\lambda + \rho} + M_j \circ \tilde{L}_i. \tag{3.8}$$

In practice, we ignore the terms with ϵ and compute them as

$$L_j = (1 - M_j) \circ \frac{\lambda\tilde{L}_i + \rho X_j}{\lambda + \rho} + M_j \circ \tilde{L}_i. \tag{3.9}$$

L can be obtained by collecting and arranging each L_j, and its total complexity is $O(nd)$. Since we do not focus on minimizing the nuclear norm or its variations, our method does not require time-consuming SVDs.

3.2.2 $\tilde{L}$-Subproblem

We first initialize $\tilde{L}$ by

$$\tilde{L}_i = \frac{1}{||p_i||_1} \sum_{(p_i)_j=1} X_j. \tag{3.10}$$

Then the $\tilde{L}$-subproblem is given by :

$$\min_{\tilde{L}} \frac{\lambda}{2} \sum_{i=1}^{c} \sum_{(p_i)_j=1} ||(1 - M_j + \epsilon) \circ (L_j - \tilde{L}_i)||_{\mathrm{F}}^2, \tag{3.11}$$

moreover, its solution is

$$\tilde{L}_i = \frac{1}{\sum_{(p_i)_j=1}(1 - M_j + \epsilon)} \circ \sum_{(p_i)_j=1} (1 - M_j + \epsilon) \circ L_j. \tag{3.12}$$

That is, with ϵ being small enough, only k with $M_{kj} = 0$ is considered in calculating $\tilde{L}_i$, except for the case when $M_{kj} = 1$ along all j with $(p_i)_j = 1$.

3.2.3 M-Subproblem

The M-subproblem can be split with respect to $j = 1, 2, ..., d$ and the corresponding i is given by

$$\begin{aligned} \min_{M} \ & \frac{\lambda}{2} \sum_{i=1}^{c} \sum_{(p_i)_j=1} ||(1 - M_j + \epsilon) \circ (L_j - \tilde{L}_i)||_{\mathrm{F}}^2 \\ & + \frac{\rho}{2} ||(1 - M) \circ (L - X)||_{\mathrm{F}}^2 \\ & + \beta ||M||_1 + \gamma \sum_{j=1}^{n} \sum_{(kj,lj) \in \mathcal{N}} |M_{kj} - M_{lj}|. \end{aligned} \tag{3.13}$$

This energy form can be solved with α-expansion algorithm with

$$\begin{aligned} D &= \frac{\lambda}{2} \sum_{i=1}^{c} \sum_{(p_i)_j=1} ||(1 - M_j + \epsilon) \circ (L_j - \tilde{L}_i)||_{\mathrm{F}}^2 \\ &\quad + \frac{\rho}{2} ||(1 - M) \circ (L - X)||_{\mathrm{F}}^2 + \beta ||M||_1, \\ V &= \gamma \sum_{j=1}^{n} \sum_{(kj,lj) \in \mathcal{N}} |M_{kj} - M_{lj}|, \end{aligned} \tag{3.14}$$

and its computational complexity per iteration is near linear according to [5]. Since (3.12) cannot be represented by a pixel-wise sum of a function of M, α-expansion cannot be applied unless we split the dynamic background from the static background.

3.2.4 p-Subproblem

The p-subproblem is given by :

$$\min_{p} \frac{\lambda}{2} \sum_{i=1}^{c} \sum_{(p_i)_j=1} ||(1 - M_j + \epsilon) \circ (L_j - \tilde{L}_i)||_{\mathrm{F}}^2. \tag{3.15}$$

Because we separated L and $\tilde{L}$, the value of each summand of (3.15) is independent of the choice of p unlike with the k-means problem. Therefore, it simply can be solved by calculating $||(1 - M_j + \epsilon) \circ (L_j - \tilde{L}_i)||_{\mathrm{F}}^2$ for every (i, j) pair and selecting i that gives the minimal value for each j.

3.2.5 Adaptive parameter control

The parameters λ and ρ control the ratio between the two terms that involve the L-subproblem. A low λ/ρ leads to L becoming similar to X in M^c. Contrarily, the larger λ/ρ, the more likely it becomes that the L-subproblem (3.7) will prefer a static L. Therefore, we set λ/ρ large at the beginning of the iterations to recover the obscured portion of L and reduce λ by a factor of s after each iteration with the lower bound being $\lambda/\rho = 0.01$. Such parameter settings leads to $L \approx X$ in M^c when the iterations are completed, which then successfully recovers dynamic backgrounds.

The parameter β controls the sparsity of M; a large β produces a sparse

M. (3.13) can be transformed into

$$\begin{aligned}\min_{M_j} \; & \sum_{k=1}^{n} \left(\beta - \frac{\lambda}{2}(L_{kj} - \tilde{L}_{ij}) - \frac{\rho}{2}(L_{kj} - X_{kj})^2 \right) M_{kj} \\ & + \gamma \sum_{(kj,lj)\in\epsilon} |M_{kj} - M_{lj}| \\ & + \frac{\lambda(1+\epsilon)}{2} ||L_j - \tilde{L}_i||_{\mathrm{F}}^2 + \frac{\rho}{2} ||L_j - X_j||_{\mathrm{F}}^2.\end{aligned} \tag{3.16}$$

From (3.16), we can see that M_{ij} is more likely to be 1 if $\beta < \frac{\lambda}{2}(L_{kj} - \tilde{L}_{ij}) + \frac{\rho}{2}(L_{kj} - X_{kj})^2$. Hence, we made β depend on the distribution of the right inequality term for M to distinguish a moving object from the background. Because the estimation of M is not accurate at the beginning of the iterations, we start with a large β and reduce it by a factor 0.5 until $||M||_1$ reaches $10^{-4}nd$ as in [69]. Thereafter, compute the variance σ^2 of $L - X$ at each iteration and set β as $\min\{\max\{0.5\beta, 4.5\rho\sigma^2, \bar{\beta}\}, \beta\}$, where $\bar{\beta}$ is the lower bound of β.

A large γ prefers a more connected L. We set $\gamma = 20\beta$, and, whenever β is updated, γ is updated with that relationship. The overall process is shown in Algorithm

2 and the source code can be found at `https://1drv.ms/u/s!AilN6GjAEcSnbAqlFP12qD-CH3Y?e=R2JyGU`.

3.2.6 Convergence

Every subproblem in the process has a converging or an analytic solution, and the total energy decreases monotonically. Therefore, the algorithm converges to a local minimum with fixed parameters. Moreover, because the parameters λ, β, γ decrease with each iteration and have a lower bound, they also converge. Therefore, the entire algorithm converges to a local minimum.

Algorithm 2 Background Subtraction by Masked RPCA

Input : $X \in \mathbb{R}^{H\times W\times 3\times d}$, $c \geq 1$, $p = \text{K-means}(X, c)$, $L = X$, $\tilde{L}$ initialized by (3.10), $M = \mathbf{0}$, λ, ρ, β, γ, s, $\tau > 0$
$X \leftarrow X_i$, where X_i is the channel of X with the largest variation.
Reshape X into an $HW \times d$ matrix
repeat
    Solve the L-subproblem by (3.9)
    Solve the $\tilde{L}$-subproblem by (3.12)
    Update β and γ according to Section 3.2.5
    Solve the M-subproblem (3.16) by using graph cuts
    $p \leftarrow \text{argmin}_p \frac{\lambda}{2} \sum_{i=1}^{c} \sum_{(p_i)_j=1} ||(1 - M_j + \epsilon) \circ (L_j - \tilde{L}_i)||_{\text{F}}^2$.
    $\lambda \leftarrow s\lambda$
until convergence or # of iterations reaches 20
$S = (|X - L| > \tau)$
Apply a 5×5 median filter on S
Output S

	Frame size	Start-end	Dynamic texture
Overpass	240 × 320 × 3	2001-3000	Waving tree
Fountain02	288 × 432 × 3	1-1000	Fountain
Canoe	240 × 320 × 3	101-1100	Flowing water
Boats	240 × 320 × 3	7000-7999	Flowing water
Camouflage	240 × 320 × 3	-	Monitor
Bootstrap	288 × 432 × 3	1-1000	Pale shadow
Foreground Aperture	240 × 320 × 3	1-1100	-
Waving Trees	240 × 320 × 3	-	Waving tree

Table 3.1: Details of the benchmark videos. "Start-end" means the frame number of the start frame and the end frame of a full sequence.

3.3 Experimental results

3.3.1 Benchmark Algorithms And Videos

We evaluated the proposed method by comparing it with other end-to-end dynamic background subtraction methods, namely, SuBSENSE[52], CANDID[35], BF-LRGB[45], and IUTIS-2[3]. In addition, Javed *et al.* [23] was included in the benchmark algorithms because they also used RPCA and MRF.

We downloaded the benchmark videos "Overpass," "Fountain02," "Canoe," and "Boats" from the CDNet dataset [59], and "Camouflage," "Bootstrap," "Foreground Aperture," and "Waving Trees" from the Wallflower dataset [55]. For CDNet, ground-truth sequences contain 5 labels : motion, unknown motion, hard shadow, static, and out-of-interest region, and we consider the first three classes as the moving object to be detected. We selected a series of 1000 frames with moving objects from each sequence and conducted experiments as in [48]. For the Wallflower dataset, when the length of the video was less than 1000, the entire clip was selected. Otherwise, the only the first 1000 frames were selected. The details of each video clip are pre-

sented in Table 3.1.

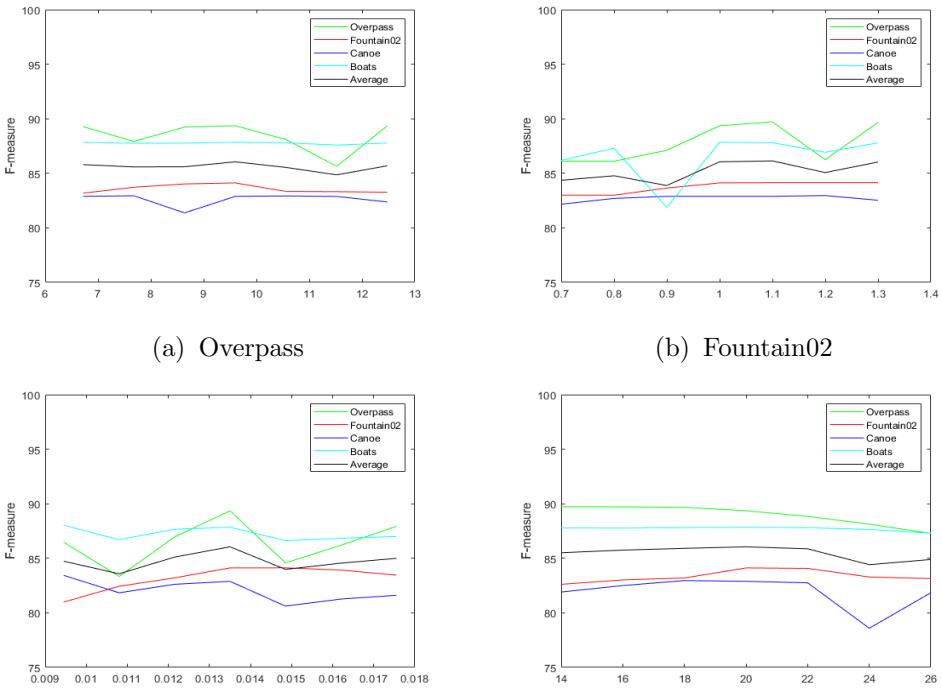
3.3.2 Implementation

We set the initial value of the parameters as $c = 1$, $\lambda = 9.6$, $\beta = 2\text{Var}(X)$, $\gamma = 20\beta$, $s = 0.7$, $\tau = 0.03$. The suggested intervals for ρ and $\bar{\beta}/\text{Var}(X)$ are [0.05, 2] and [0.01, 0.1] respectively, where $\text{Var}(\cdot)$ means the variation function. We chose $\rho = 1$ and $\bar{\beta}/\text{Var}(X) = 0.0135$. They were adaptively controlled according to Section 3.2.5. Parameter sensitivity for some important variables are shown in Figure 3.2. Computational complexity is measured by resizing the input videos. We resized CDNet videos by increasing the resize factor from 0.1 to 1 by 0.1 and recorded per-iteration time cost. Since the number of the pixel is proportional to the square of the resize factor, we compared the time cost graph with the graph of $y = ax^2$ where a is set to both of the graphs coincides at $x = 0.1$. Figure 3.3 shows the comparison result, and one can empirically verify that the per-iteration computational complexity is near linear to the number of the entire number of pixels.

3.3.3 Evaluation

Quantitative Evaluation For each video, the quantitative evaluation was performed on the basis of the precision, recall, and F-measure. The results of quantitatively evaluating our algorithm against the other benchmark algorithms are shown in Table 3.2 and 3.3. Because Javed *et al.* did not offer the source code and didn't even specify the values of the parameters in the thesis, we fine-tuned it for each input video and found the best-performing parameters.

For the CDNet data, Javed *et al.* produced few false positive(FP) pixels but also few true positives (TP) pixels thus exhibiting a low F-measure. BF-LRGB demonstrated the second-best recall in "Overpass," "Canoe," and



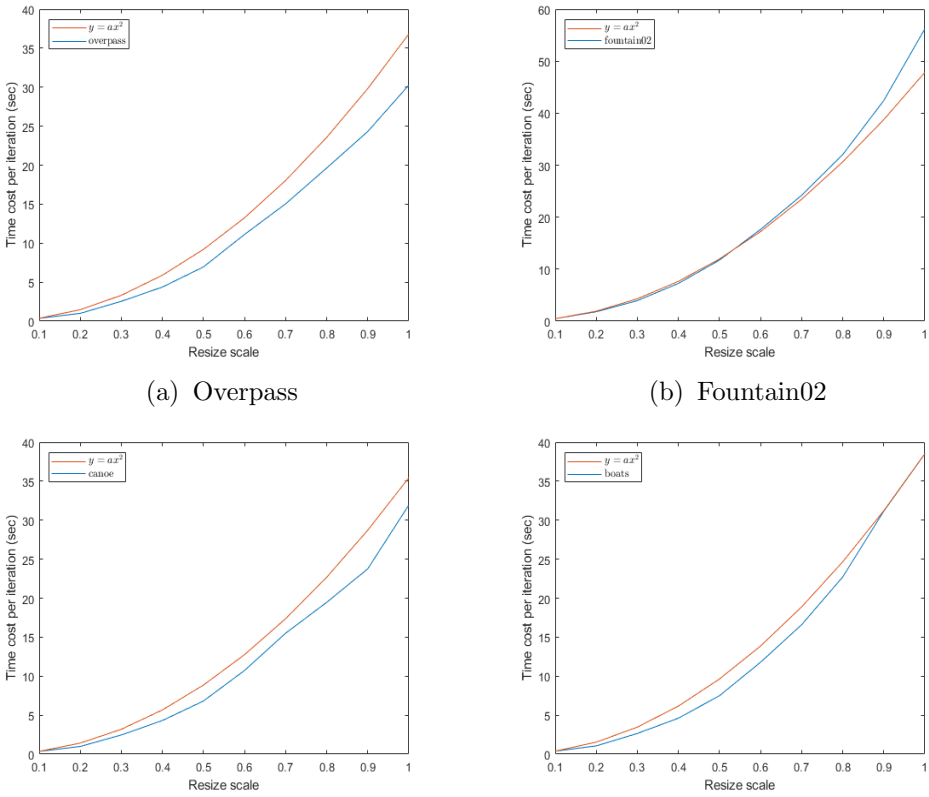
(a) Overpass

(b) Fountain02

(c) $\bar{\beta}/\mathrm{Var}(X)$

(d) γ/β

Figure 3.2: Parameter sensitivity of λ, ρ, β, and γ. Because β usually decreases to $\bar{\beta}$, we measured the sensitivity of $\bar{\beta}$ rather than that of the initial β. The range of each parameter is -30 to 30% of the value set in Section 3.3.2.



(a) Overpass (b) Fountain02

(c) Canoe (d) Boats

Figure 3.3: Per-iteration time consumption of the proposed method. One can empirically verify that the per-iteration computational complexity is near linear to the number of the entire number of pixels.

"Boats", but it produced too many FP pixels in "Overpass." Sample-based methods, namely, CANDID and SuBSENSE produced few FP pixels in all the CDNet videos. For precision, they took first and second place for all the video tests except "Boats." However, they produced a large number of false negative pixels in "Canoe," and "Boats". IUTIS-2 also showed higher precision than recall but demonstrated higher recall than the sample-based methods. The proposed method delivered balanced results. It exhibited the best performance in terms of recall and exceeded the average F-measures of the other algorithms by more than 9%.

For the Wallflower data, BF-LRGB demonstrated high precision and, in total, was the second-best by F-measure. The proposed method was the best in "Camouflage", the second-best in "Bootstrap" and "Foreground Aperture", but was the best overall.

Qualitative Evaluation The background subtraction results are shown in Figure 3.4, 3.5, and 3.6. The white region is true positive(TP), black region is true negative (TN), green region is false positive(FP), red region is false negative(FP), and gray region is the out-of-interest region.

The proposed method detected the lower body in "Overpass" most successfully. Moreover, it was the only algorithm that successfully detected the rear portion of the car in "Fountain02", and the bottom of the boats in "Boats." The sample-based methods failed to detect the right side of the body in "Foreground Aperture". In particular, Javed *et al.* also failed to detect the right side of the body because it computes its low-rank output in an online manner without considering the entire video. That is, although it follows the RPCA framework, it shares the disadvantages of the sample-based methods by identifying temporarily stationary objects as part of the background.

[%]		Javed *et al.*	CANDID	BF-LRGB	SuBSENSE	IUTIS-2	Proposed method
Overpass	Precision	89.66	**99.37**	45.74	*96.38*	95.57	90.71
	Recall	55.32	65.93	*83.76*	68.07	76.65	**88.04**
	F-measure	68.43	79.27	59.17	79.79	*85.07*	**89.36**
Fountain02	Precision	74.13	**98.97**	62.16	*97.10*	91.54	88.77
	Recall	31.06	52.11	55.17	75.15	*79.79*	**79.93**
	F-measure	43.78	68.27	58.46	*84.72*	**85.26**	84.12
Canoe	Precision	98.51	*99.05*	74.42	**99.43**	84.94	81.82
	Recall	57.88	53.34	*74.15*	52.64	64.94	**83.99**
	F-measure	72.92	69.34	*74.29*	68.84	73.61	**82.89**
Boats	Precision	91.41	**95.95**	85.61	94.29	*94.41*	87.09
	Recall	60.04	32.33	*72.17*	41.83	46.76	**88.62**
	F-measure	72.48	48.36	*78.32*	57.95	62.55	**87.85**
Average	F-measure	64.41	66.31	67.56	72.82	*76.62*	**86.05**

Table 3.2: Quantitative evaluation on CDNet dataset. For each sequence, the best result is highlighted in **bold** and the second-best in *italics*.

[%]		Javed *et al.*	CANDID	BF-LRGB	SuBSENSE	IUTIS-2	Proposed method
Camouflage	Precision	**96.55**	87.24	94.88	95.21	93.29	*95.66*
	Recall	98.37	98.81	**99.94**	97.52	98.40	*99.69*
	F-measure	*97.45*	92.67	97.34	96.35	95.78	**97.63**
Bootstrap	Precision	75.55	71.80	54.57	**89.65**	*85.65*	78.53
	Recall	42.78	27.92	**94.68**	27.81	*69.70*	63.35
	F-measure	54.62	40.21	69.24	42.45	**76.86**	*70.13*
Foreground Aperture	Precision	**92.38**	49.04	87.95	81.29	57.15	*89.91*
	Recall	69.48	54.99	**99.01**	57.46	63.07	*90.56*
	F-measure	79.31	51.85	**93.15**	67.33	59.97	*90.23*
Waving Trees	Precision	88.22	94.52	78.71	**96.67**	88.97	*96.34*
	Recall	78.39	77.78	*99.59*	94.99	**99.88**	87.37
	F-measure	83.01	85.34	87.93	**95.83**	*94.11*	91.63
Average	F-measure	78.60	67.51	*86.92*	75.49	81.68	**87.41**

Table 3.3: Quantitative evaluation on Wallflower dataset. For each sequence, the best result is highlighted in **bold** and the second-best in *italics*.

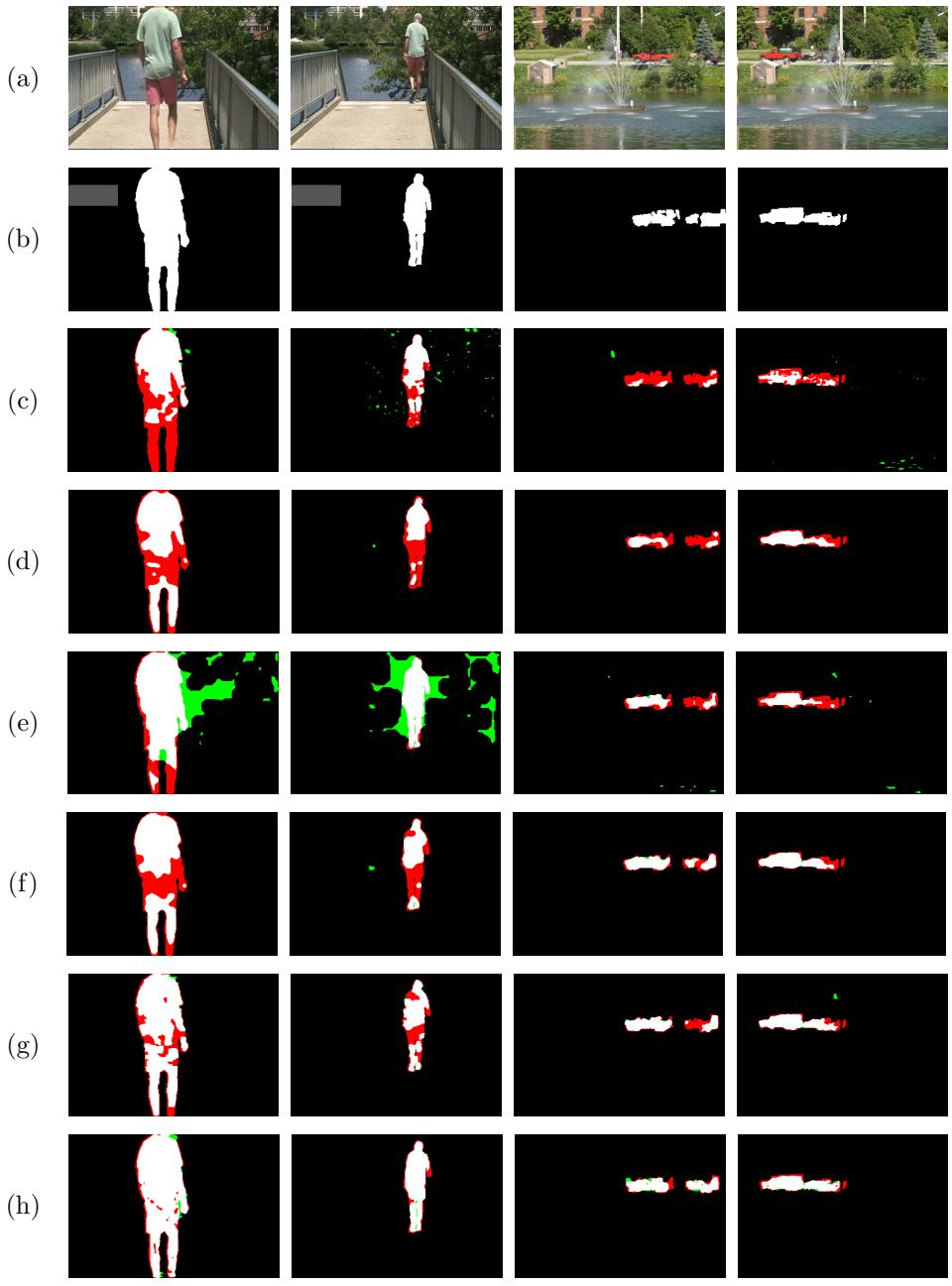


Figure 3.4: Background subtraction results. (a) Input frame, (b) ground truth, and the experimental results of (c) Javed *et al.*, (d) CANDID, (e) BF-LRGB, (f) SuBSENSE, (g) IUTIS-2, and (h) the proposed method applied to Overpass (column 1 and 2) and Fountain02 (column 3 and 4).

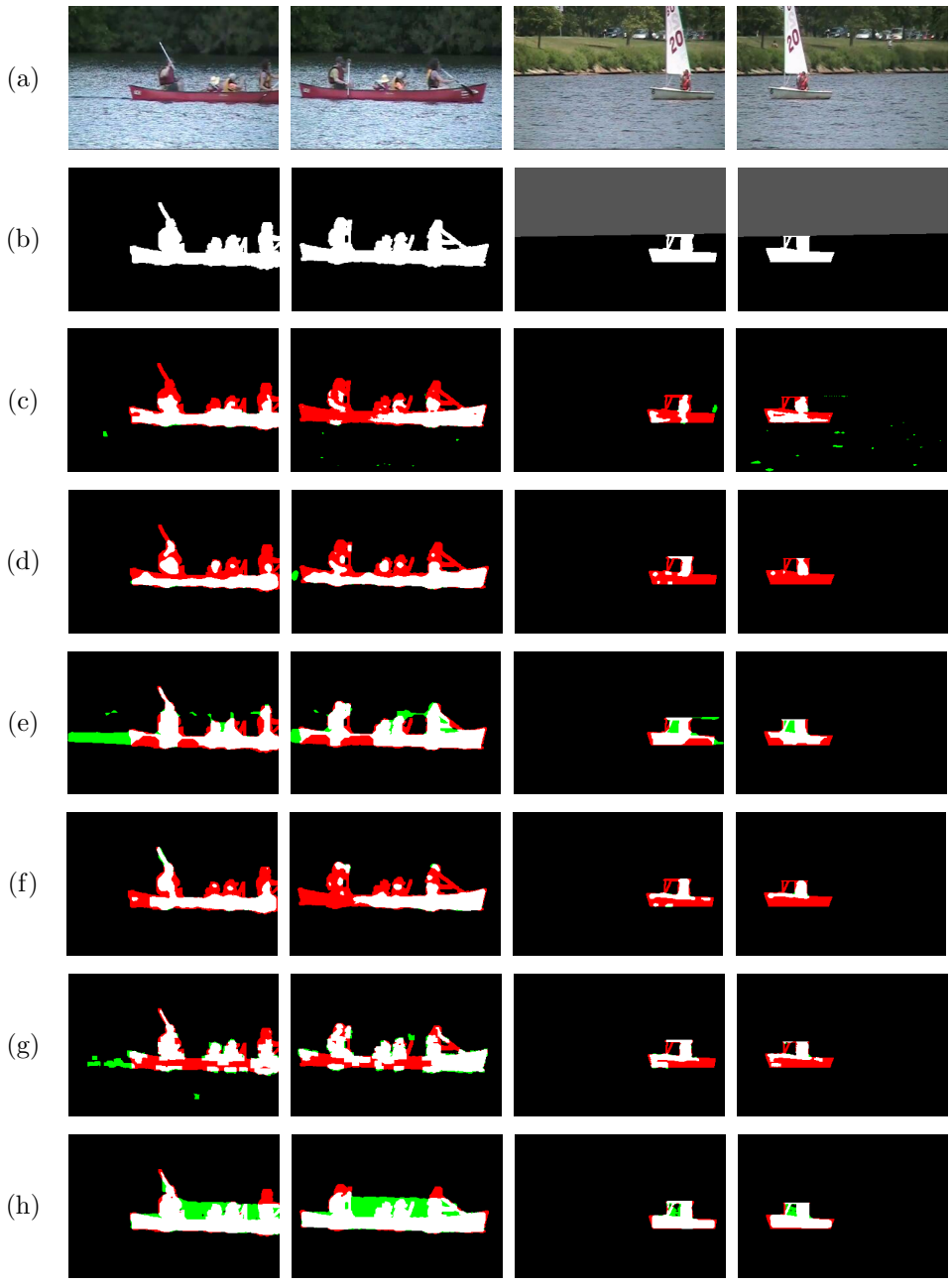


Figure 3.5: Background subtraction results. (a) Input frame, (b) ground truth, and the experimental results of (c) Javed *et al.*, (d) CANDID, (e) BF-LRGB, (f) SuBSENSE, (g) IUTIS-2, and (h) the proposed method applied to Canoe (column 1 and 2) and Boats (column 3 and 4).

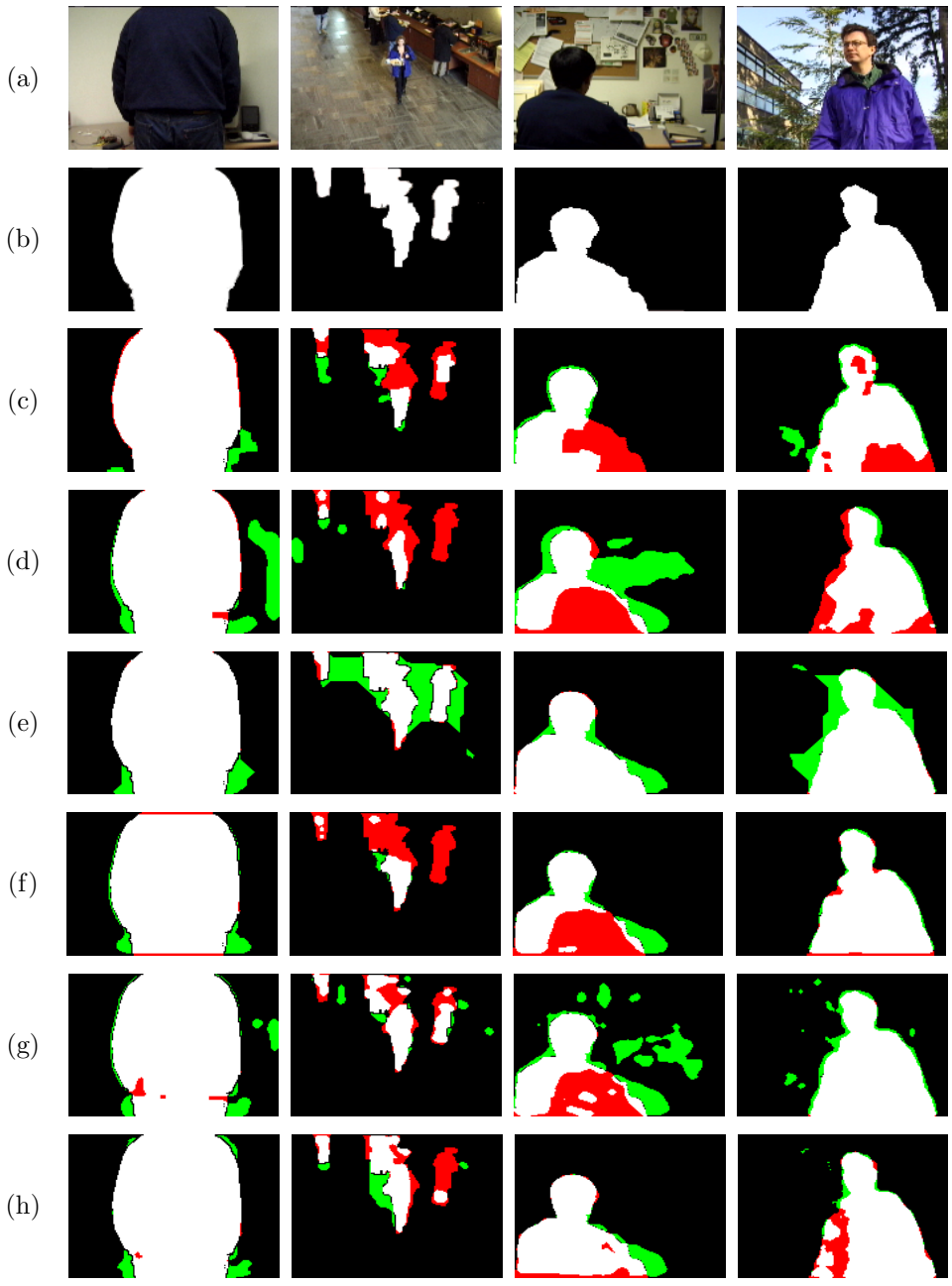


Figure 3.6: (a) Input frame, (b) ground truth, and the experimental results of (c) Javed *et al.*, (d) CANDID, (e) BF-LRGB, (f) SuBSENSE, (g) IUTIS-2, and (h) the proposed method applied to Camouflage (column 1), Bootstrap(column 2), Foreground Aperture (column 3) and Waving Trees (column 4).

Chapter 4

Deep Image Dedusting With Dust-Optimized Transmission Map

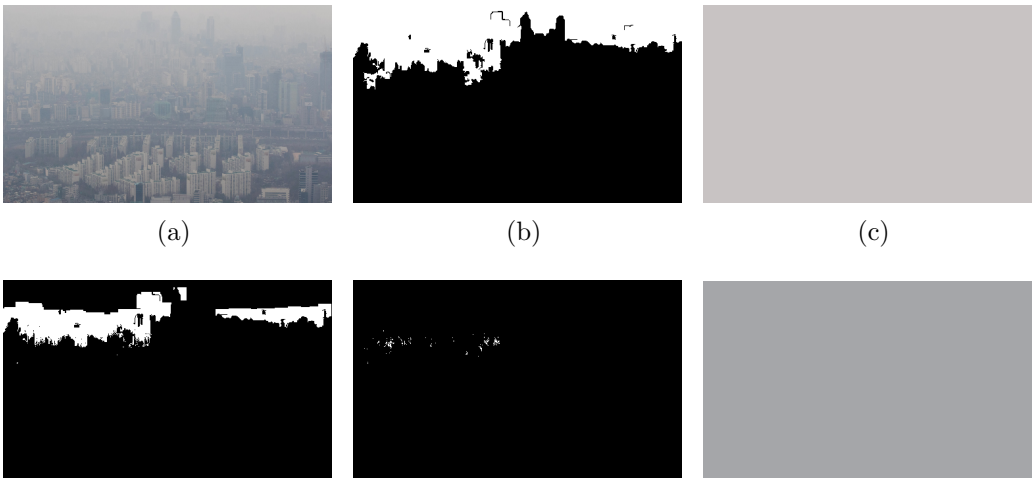
4.1 Transmission estimation

4.1.1 Atmospheric light estimation

Most existing dehazing techniques use the atmospheric light estimating method developed by [17]. That is, when 0.1% of pixels with the largest dark channel values are denoted as D, the value of the brightest pixel in D is taken as atmospheric light. Denote atmospheric light computed in this way as A_{old}. As sunlight is so bright, this method gives A as the mixture of the light from the dense haze and the unscattered sunlight, which is brighter than the light reflected from the densest haze. Since many t estimating methods

assume that A is given, this implies that sunlight influences the selection of t, which violates (2.34). In the case of dehazing problem, this error is not a big problem because haze has high intensity. However, as dust is much darker than haze, this error induces an overestimation of A, which results in too dark and unnatural dedusting results. To find darker atmospheric light and address this problem, we must find the dustiest region. Since finding this region does not need to be too precise, we used Felsenszwalb's efficient graph-based image segmentation [12]. The region of interest is the component containing most of D among the segmented components and is denoted as ROI.

Since most of the dusty region is usually located in the sky region of images, a component in the sky region is often selected as the region of interest. In ROI, we propose the following atmospheric light estimation method:



(a) (b) (c)

(d) (e) (f)

Figure 4.1: (a) Dusty image. (b) ROI selected by segmentation. (c) A_{old}. (d) 50% pixels with smallest dark channel values in ROI. (e) Remaining pixels after filtering with intensity and saturation values. (f) A_{new}. A_{new} is darker and less saturated than A_{old}.

First, choose the 50% of pixels with the smallest DC_{ROI} values in ROI. Secondly, choose 10% of pixels with the smallest intensity values in the remaining area. Finally, choose 10% of pixels with the smallest saturation values in the remaining area, and the average value of them is taken as atmospheric light. Atmospheric light found using our method is darker and less saturated than the classical estimation A_{old}. We call this estimation A_{new}. Figure 4.1 shows A_{old} and the process of finding A_{new}.

4.1.2 Transmission estimation

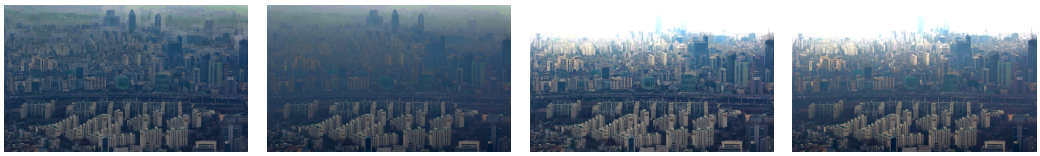
The initial estimation of the transmission map based on dark channel prior is given by [17]

$$t_{DC}(x) = 1 - \min_{y \in \Omega(x)} \min_{c} \frac{I^c(y)}{A^c}. \tag{4.1}$$

It is obtained by assuming constant t in $\Omega(x)$, dividing (2.35) by A, and applying (2.34). This estimation reflects distance correctly for most regions but causes a halo effect since its patch-wise calculation overestimates the transmission map at the boundary of front objects. The ways to overcome this are soft-matting [26], guided filtering [16], another initial estimation based on the boundary constraint [36], and so on, each having its pros and cons in terms of accuracy and speed. We chose soft-matting which is the slowest but most accurate. The transmission map refined by soft-matting is denoted as t_{SM}, and can be obtained with following equation:

$$(L + 10^{-4}U)t_{SM} = 10^{-4}t_{DC} \tag{4.2}$$

where L is the Laplacian matting matrix, U is the identity matrix, and t_{DC} is flattened. Depending on which A is used in calculating t_{DC}, *old* or *new*



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

Figure 4.2: (a) $J(t_{DC}^{old}, A_{old})$, (b) $J(t_{SM}^{old}, A_{old})$, (c) $J(t_{DC}^{new}, A_{new})$, and (d) $J(t_{SM}^{new}, A_{new})$. Ones calculated with A_{old} not only look dark, but also show awkward and unrealistic hue. The sky region of ones calculated with A_{new} has been whitened and lost information of there. Regardless of A, one can see that recovery with t_{DC} shows a halo effect but shows higher local contrast in the front buildings than recovery with t_{SM}.

will be displayed in the superscript of t_{DC} and t_{SM}.

Let J_{rough} be a rough recovery of scene radiance defined by

$$J_{rough}(t, A) = \frac{I - A}{\max\{t, 0.1\}} + A. \tag{4.3}$$

Figure 4.2 shows $J(t_{DC}^{old}, A_{old})$, $J(t_{SM}^{old}, A_{old})$, $J(t_{DC}^{new}, A_{new})$, and $J(t_{SM}^{new}, A_{new})$.

Roughly, A_{new} is the average of the darkest pixels of the sky. Also, roughly, t_{DC} and t_{SM} gives small values to regions whose value is close to A. This makes t_{DC}^{new} and t_{SM}^{new} have zero value at highly dusty regions, which rarely occurs with A_{old}. In short, transmission with A_{new} gives good estimation at close and less dusty regions, and transmission with A_{old} gives the right intensity ordering of transmission. Since we assumed (2.33), t_{SM}^{old} and t_{SM}^{new} should differ by constant exponent. Therefore, to complement each other, we find

$$f(x) = x^a \tag{4.4}$$

(a) (b) (c)

(d) (e) (f)

Figure 4.3: (a) t_{DC}^{old}, (b) t_{SM}^{old}, (c) graph of f and t_{SM}^{new} over flattened and sorted t_{SM}^{old}, (d) t_{DC}^{new} (e) t_{SM}^{new}, and (f) t_0 .

where a minimizes

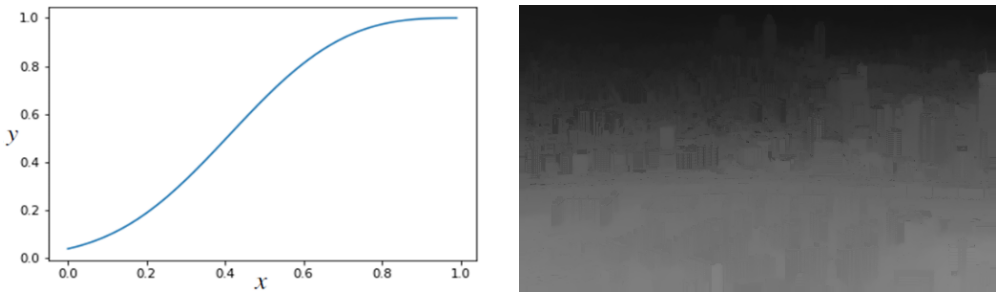
$$||f(\Omega_{old}) - \Omega_{new}||_2 \tag{4.5}$$

where Ω_{old} and Ω_{new} are the 20% of pixels with the largest transmission values in ROI^c of t_{SM}^{old} and t_{SM}^{new}, flattened and rearranged in ascending order. We finally define our initial estimation of the transmission t_0 by

$$t_0 = f(t_{SM}^{old}). \tag{4.6}$$

Then t_0 is similar to t_{SM}^{new} at close, high transmission regions but higher than t_{SM}^{new} at low transmission regions.. Also, its lightness order among pixels is identical to that of t_{SM}^{old}. Fig. 4.3 shows the process of generating t_0.

The transmission map should be smooth except at depth discontinuities.



(a) (b)

Figure 4.4: (a) Graph of $y = \exp\left(-\beta\left(1-x\right)^3\right)$ with $\beta = 1.309$, which is calculated by (4.12), (b) t which is the minimizer of (4.7). In (b), one can see that the undesirable details in the front have largely disappeared while the back remains well bordered.

Although soft-matting successfully removes over-estimation of transmission and corresponding halo effects, it induces underestimation of transmission at near and dark region. Such underestimations result in over-enhancement, which distorts the original scene radiance. Therefore, a refinement operation to remove them should be applied to t_0. To remove such undesirable discontinuities and preserve semantic edges, NLTV regularization is applied to refine t_0 as [33]. Since the damage of over-enhancement at close region is more critical than at far regions, distance adaptive regularization is needed. Therefore, we minimize the following objective function:

$$\min_t \sum_x u(x)|\nabla_w t(x)| + \frac{\lambda}{2}||t - t_0||_{\mathrm{F}}^2, \tag{4.7}$$

where u is a weight function, $\nabla_w t(x)$ is the non-local gradient of t at x, and λ is a weight to balance two terms. The initial estimation of the transmission

map t_0 is used to reflect distance, so we can define weight u by

$$u(x) = \exp\left(-\beta\left(1 - t_0(x)\right)^3\right). \tag{4.8}$$

We denote the minimizer of (4.7) as t, which is the final estimation of transmission. t is shown in Figure 4.4.

4.2 Scene radiance recovery

Once we have t, the scene radiance may be recovered according to (2.37) as

$$J = \max\left(\min\left(\frac{A}{1-\omega \log t} + \frac{I-A}{t(1-\omega \log t)}, 1\right), 0\right), \tag{4.9}$$

where $\omega = p(0,0)/4\pi$. However, dividing by $t(1+\omega \log t))$ not only amplifies noise but also whitens the bright part of the sky whose transmission is close to zero, but $I - A$ is not small enough. It is because the dark channel prior is not sufficiently reliable at the sky region, and this problem is evaded by alleviating the dust model (2.37) with setting a lower bound of coefficient of J as s:

$$I_{J,s,\omega}(x) = J(x)\max\left(s, t(x)\left[1 - \omega \log t(x)\right]\right) + A(1 - t(x)). \tag{4.10}$$

In this case, choosing optimal s and ω requires additional human-made priors. The lower bound of t is often set 0.1 in the field of dehazing [17], [33]. However, since we have to find ω in addition to s, using fixed s is not desirable. [60] chose to maximize the number of visible edges, but it is too unstable that it often fails to attain naturalness. Instead of searching for other human-crafted regularizers to fit (4.9), we made use of deep image

prior(DIP) [56]. DIP is a neural-based image prior, which replaces image variable in regularizers with the output of a generative neural network with random noise inputs. It is based on the idea that the structure of a generator network is sufficient to capture low-level image information, and it showed great success in many variational models, e.g. denoising, super-resolution, inpainting, and deblurring [61] without training. [61] solved blind deconvolution problem by constructing two generator G_X and G_k, where G_X generates deblurred images and G_k generates blur kernels. Likewise, we propose constructing three generators G_J, G_s, and G_ω, each generating J, s, and ω. G_J is an asymmetric autoencoder with skip connections followed by the Sigmoid activation. G_J has five encoding modules and corresponding five decoding modules, and their detailed structures are shown in Table 4.1, Figure 4.5 and Figure 4.6. Since G_s and G_ω process much simpler process than G_J, we adopted fully connected networks. They take length 200 random vectors as inputs, and the outputs are generated through one hidden layer of 2000 nodes. Finally, the parameters of G_J, G_s, and G_ω are optimized by simply minimizing following mean square error loss function:

$$\mathcal{L}(G_J, G_s, G_\omega) = ||I_{J,s,\omega} - I||_{\mathrm{F}}^2. \tag{4.11}$$

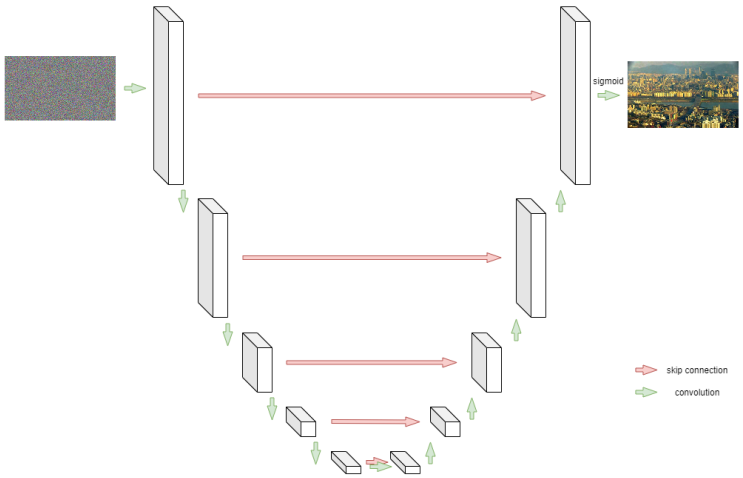


Figure 4.5: Architecture of G_J. Left five block are encoders e_1 to e_5, and right five blocks are decoders d_1 to d_5. Downwards arrows accompany a stride 2 convolution which downsizes the feature map by the factor of 2. Upwards arrows upscale the feature map with bilinear interpolation. Skip connection is conducted by concatenating in the feature dimension.

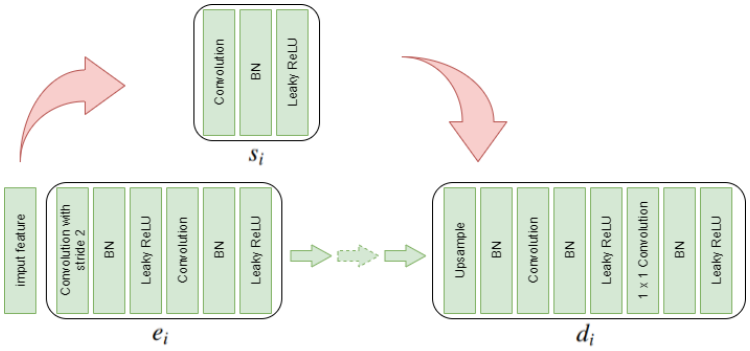


Figure 4.6: Architecture of each e_i, s_i, and d_i. For $i \geq 3$, a non-local block is added after the first leaky ReLU of e_i. The slope of negative part of leaky ReLU is 0.2. Concatenation (pink arrow) is performed right after upsampling.

Input : z_x $(h \times w \times 3)$ from the uniform distribution with seed 0.	
Output : Dedusted image x $(h \times w \times 3)$.	
Encoder 1	$e_1(128, 3),\ s_1(16, 3)$
Encoder 2	$e_2(128, 3),\ s_2(16, 3)$
Encoder 3	$e_3(128, 3),\ s_3(16, 3)$
Encoder 4	$e_4(128, 3),\ s_4(16, 3)$
Encoder 5	$e_5(128, 3),\ s_5(16, 3)$
Decoder 1	$d_1(128, 3)$
Decoder 2	$d_2(128, 3)$
Decoder 3	$d_3(128, 3)$
Decoder 4	$d_4(128, 3)$
Decoder 5	$d_5(128, 3)$
Output layer	Conv(1, 1) & Sigmoid

Table 4.1: Structure of each encoder/decoder module. For $m \in \{e, s, d, \text{Conv}\}$, $m(n, k)$ represents that each convolution of m has n $k \times k$ kernels. All convolutions are done with appropriate zero padding that maintains the feature shape.

4.3 Experimental results

4.3.1 Implementation

Because soft matting and NLTV regularization takes tremendous time, all processes prior to scene radiance recovery are done by resizing the shorter edge of the input image to a length of 150. After we have t, we upsample it into original shape. In the segmentation step, to select ROI, I is converted to a grayscale image and smoothed by a Gaussian filter with standard deviation 0.8 as the input. The scale parameter k is set to 50. As k becomes larger, $|ROI|$ gets larger. Patch size for dark channel calculation is set to 21×21.

β depends on the mean and the standard deviation of dark channel values at ROI^c. β is set to

$$\beta = -\frac{\log 0.4}{(1 - \text{mean}\left[t^{new}_{DC_{ROI^c}}\right] + \text{std}\left[t^{new}_{DC_{ROI^c}}\right])^2}, \tag{4.12}$$

where mean[$\cdot$] and std[$\cdot$] are the mean and the standard deviation of $\cdot$, respectively.

For NLTV regularization (4.7), We set λ and μ as 1 and 20, respectively. The patch size for NLTV calculation is set to $11\times$ 11, 10 5 $\times$5 searching window is employed for each area $A(x)$, and σ at 2.38 is set to 255×10^{-2}. (4.7) is solved as follows; flatten t_0, introduce an auxiliary variable d which approximates $\nabla_w t(x)$, split d and $\nabla_w t(x)$, find N such that $\nabla_w t(x) = N \cdot t$, and alternatively solve d and t from

$$\min_{t,d} \sum u(x)|d(x)| + \frac{\lambda}{2}||t - t_0||_2^2 + \frac{\mu}{2}||d - N \cdot t||_2^2. \tag{4.13}$$

With t fixed, d is updated first by applying the shrinkage operator:

$$d^{k+1}(x) = \frac{\nabla_w t^k(x)}{|\nabla_w t^k(x)|} \max\left(|\nabla_w t^k(x)| - \frac{u(x)}{\mu}, 0\right). \tag{4.14}$$

Then, t is updated by solving following sparse linear system

$$\left(\frac{\lambda}{\mu}E + N^{\mathrm{T}}N\right) t^{k+1} = \frac{\lambda}{\mu}t_0 + N^{\mathrm{T}}d^{k+1}, \tag{4.15}$$

where $E \in \mathbb{R}^{hw \times hw}$ is the identity matrix. We found that only one iteration is enough for good results.

(4.11) is minimized by Adam optimizer. The learning rates are set 10^{-2} for G_J and 10^{-4} for G_s and G_ω, and they decay by the factor of 0.5 for 2000, 3000, and 4000 iterations.

4.3.2 Evaluation

We compared our method with state-of-art dehazing methods, namely, Cycle-dehaze [11], DHSGAN [34], and DCPDN [66]. We also included Wang et al.[60] in the benchmark algorithms since it is the unique existing dedusting algorithm. Because the standard dedusting dataset does not exist, we selected 8 outdoor dusty images from Google's image. Also, because there is no clean reference image and dedusting assessment algorithm, quantitative evaluation is performed on the basis of no-reference image quality assessment methods used in related fields. We used MDM [37], CEIQ [63], LOE [58], UCIQE [64], and α_{SB} and β_{SB} introduced in [46]. MDM, CEIQ, and LOE focus on contrast enhancement. MDM is based on higher orders of Minkowski

distance, defined by

$$\mathrm{MDM(J)} = \sqrt[4]{\left(\frac{1}{N}\sum_{i=1}^{N}|J_i^8 - \mathrm{mean}(J^8)|^{64}\right)^{1/64}}, \tag{4.16}$$

where i indicates each pixel and N is the number of pixels of J. CEIQ measures histogram based entropy and cross entropy between the original image and the enhanced one. LOE assesses the naturalness of the enhanced image by measuring how well the lightness order is preserved. Unlike the other metrics, lower LOE is better. UCIQE is a underwater color image quality evaluation metric. α_{SB} measures how small the dark channel value is compared to the input image. β_{SB} is based on haze line theory [2], and the explanation for it is shown in Figure 4.7.

Dusty image enhancements are shown in Figure 4.8 and 4.9. It is the dedusting result of images 1,2,3,4 from the left column of Figure 4.8, and the result of images 5,6,7,8 from the left column of Figure 4.9. Wang et al. and Cycle-Dehaze severely distorted the color and style of the image. Wang et al. optimized the parameters through the genetic algorithm using the gradient calculated with the Kirsch operator as the evaluation function. The value of that gradient is higher than that of the result of the proposed method, but the naturalness of the image is largely lost. Both DHSGAN and DCPDN usually produced less enhanced results. In image 3 4.8, the boundary between the mountains and the sky at the top of the image is more visible in the proposed method than that of DHSGAN and DCPDN. This is also true of the boundary between the mountain and the sky in the image 5. In image 6, DHSGAN remained dust at the top of the image and it seems like clouds. In image 7, DCPDN failed to remove dust on the middle right of the picture, leaving a white band. In image 8, both DCPDN and DHSGAN made white dust bands in the middle of the image while the proposed method successfully dedusted the image.

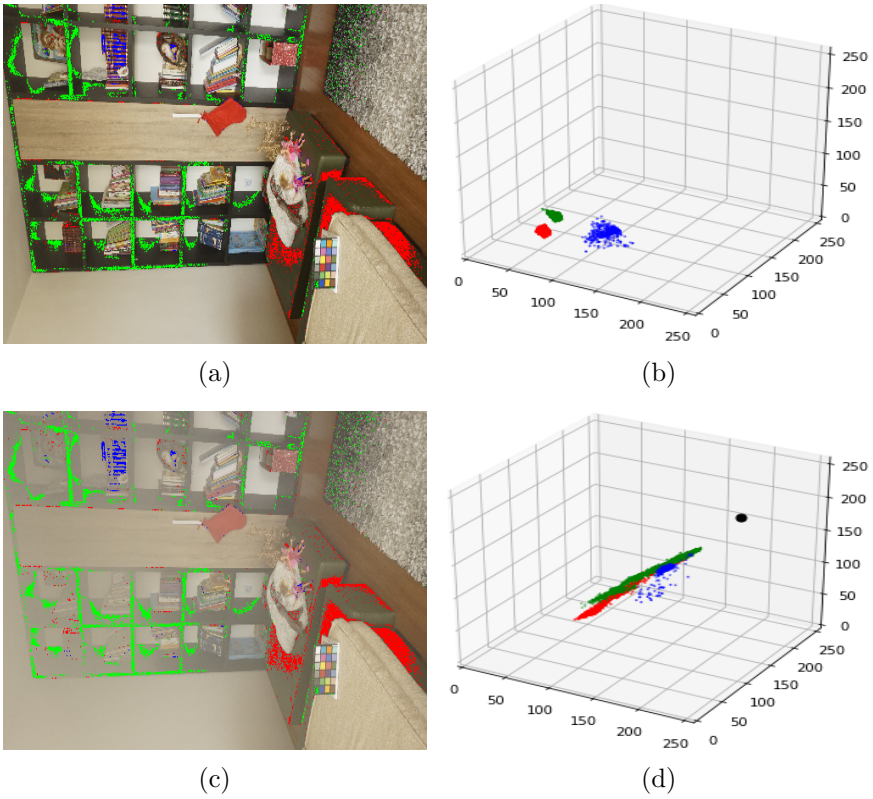


Figure 4.7: (a), (b) Pixel value of haze free natural images forms tight clusters in RGB space. (c), (d) When scattering obstructs the scene, these pixels translate according to haziness in that region and form haze lines that pass through atmospheric light A marked as the black point. β_{SB} evaluates how small the standard deviation of the magnitude of each haze line is compared to that of the input image.

Since Wang et al. and Cycle-Dehaze produced unrealistic results, quantitative evaluations were carried out only on DHSGAN, DCPDN and the proposed method. The quantitative evaluation result is shown in Table 4.2. The proposed method gave the best result for most of the scores.

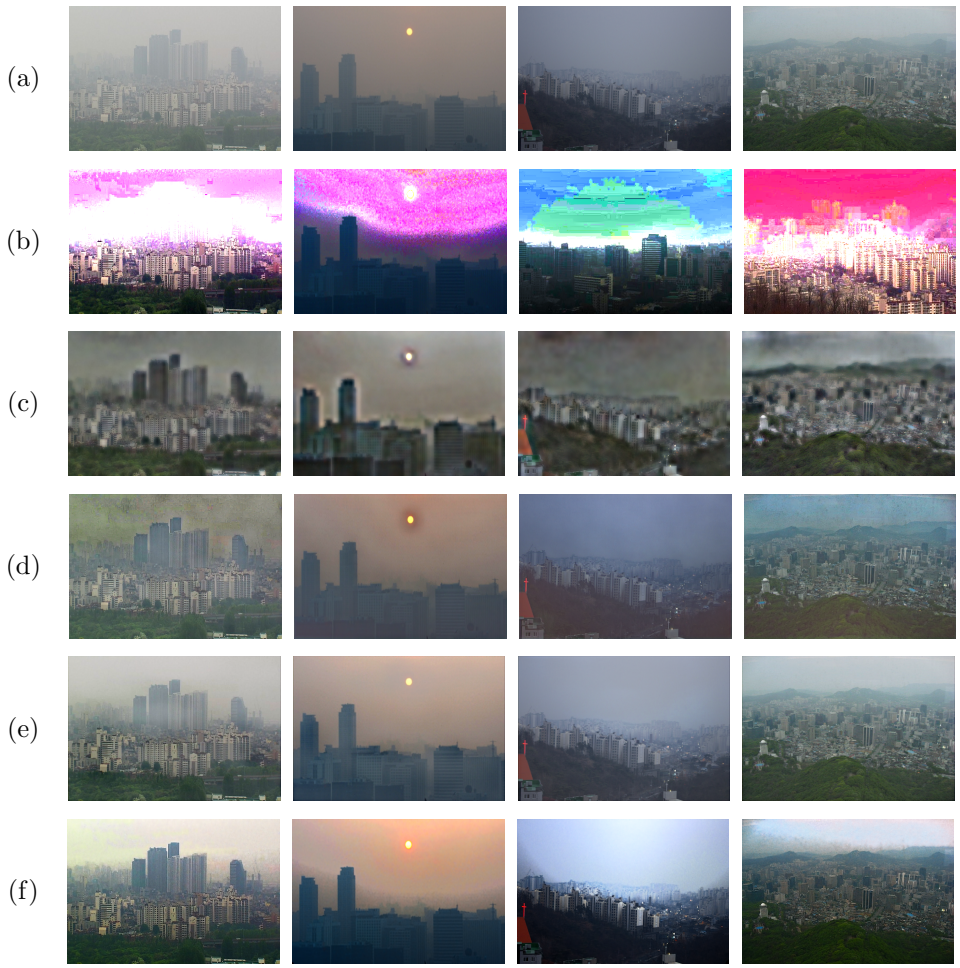


Figure 4.8: Dedusting results. (a) Input image, result of (b) Wang et al. [60], (c) Cycle-dehaze [11], (d) DHSGAN [34], (e) DCPDN [66], and (f) the proposed method applied to of image 1 to 4.

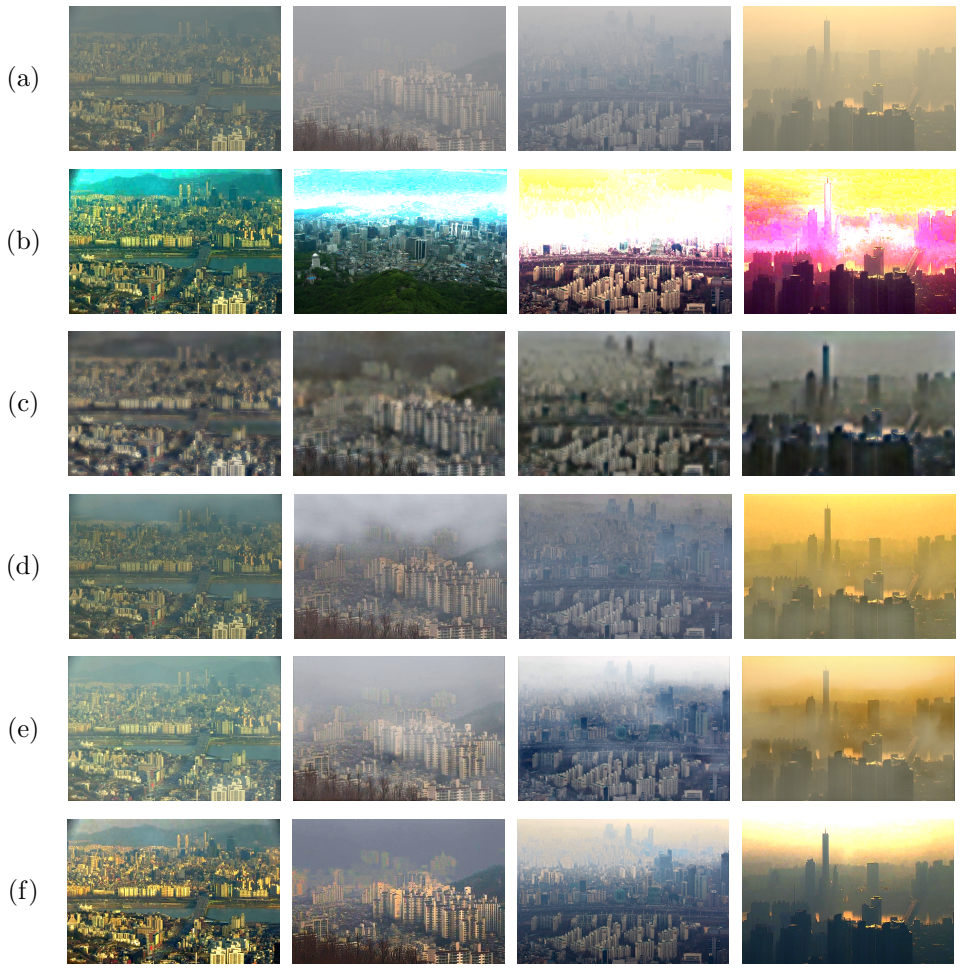


Figure 4.9: Dedusting results. (a) Input image, result of (b) Wang et al. [60], (c) Cycle-dehaze [11], (d) DHSGAN [34], (e) DCPDN [66], and (f) the proposed method applied to of image 5 to 8.

		[34]	[66]	Ours			[34]	[66]	Ours
img 1	MDM	0.6655	0.6112	**0.8917**	img 5	MDM	0.8547	0.9546	**0.9668**
	CEIQ	2.7787	3.2608	**3.4739**		CEIQ	2.8549	2.9374	**3.5355**
	LOE	1888.5	559.8	**386.5**		LOE	1240.9	739.4	**576.3**
	UCIQE	16.973	17.453	**29.563**		UCIQE	15.902	14.629	**31.548**
	α_{SB}	0.0598	0.0315	**0.0795**		α_{SB}	0.0091	0.0030	**0.0490**
	β_{SB}	0.0027	0.0031	**0.0146**		β_{SB}	0.0016	0.0012	**0.0078**
img 2	MDM	**0.9700**	0.9563	0.9632	img 6	MDM	0.7165	0.9092	**0.9562**
	CEIQ	2.8280	3.2786	**3.3102**		CEIQ	**3.1212**	3.0975	3.0643
	LOE	508.4	416.9	**187.1**		LOE	1331.1	**601.5**	932.2
	UCIQE	15.201	20.533	**29.834**		UCIQE	14.827	19.760	**27.818**
	α_{SB}	0.0030	0.0036	**0.0133**		α_{SB}	0.0208	0.0091	**0.0681**
	β_{SB}	0.0009	0.0007	**0.0021**		β_{SB}	0.0052	0.0012	**0.0059**
img 3	MDM	**0.9573**	0.9146	0.9371	img 7	MDM	0.5189	0.9200	**0.9515**
	CEIQ	2.8231	3.1322	**3.4461**		CEIQ	2.4884	3.4078	**3.531**
	LOE	444.3	491.2	**325.2**		LOE	2211.7	994.9	**322.6**
	UCIQE	15.292	24.673	**29.389**		UCIQE	12.859	21.876	**23.408**
	α_{SB}	0.0067	0.0052	**0.0318**		α_{SB}	0.0331	0.0460	**0.0524**
	β_{SB}	0.0010	0.0004	**0.0170**		β_{SB}	0.0009	0.0075	**0.0116**
img 4	MDM	**0.9543**	0.9180	0.9182	img 8	MDM	**0.9532**	0.932	0.9396
	CEIQ	2.9085	3.3147	**3.5355**		CEIQ	3.0664	3.0864	**3.4733**
	LOE	731.2	**314.6**	446.9		LOE	713.8	753.4	**478.2**
	UCIQE	15.403	22.758	**26.623**		UCIQE	17.488	14.499	**30.513**
	α_{SB}	0.0224	0.0016	**0.0226**		α_{SB}	0.0497	0.0385	**0.0615**
	β_{SB}	0.0019	0.0002	**0.0043**		β_{SB}	0.0003	0.0015	**0.0030**

Table 4.2: Quantitative evaluation of DHSGAN[34], DCPDN [66], and the proposed method. Except for LOE, the larger value is the better. The best result is written in **bold**.

Chapter 5

Conclusion

In this thesis, we proposed a new dynamic background subtraction method termed masked RPCA. The proposed method estimates the background by minimizing the multiplication of the rank term and the background mask $1 - M + \epsilon$. As the iteration proceeds, the parameters are adaptively updated to refine M and L. The per-iteration computational complexity of the proposed method is near linear. Although our method did not produce in every aspect the best quantitative result compared to the state-of-art model, it exceeded the average F-measure of the other algorithms for both the CDNet and Wallflower data due to balanced high precision and recall.

Subsequently, we proposed an effective framework to estimate appropriate atmospheric light and transmission for the dedusting problem using segmentation and dark channel prior. For transmission refinement, NLTV regularization with respect to dark channel was used. To recover the scene radiance, deep image prior was applied and corresponding generative networks are designed. Experiments showed that the proposed method outperforms the state-of-art dehazing methods, which prove its value as a proprietary algorithm for dedusting.

Bibliography

[1] Olivier Barnich and Marc Van Droogenbroeck. Vibe: A universal background subtraction algorithm for video sequences. *IEEE Transactions on Image processing*, 20(6):1709–1724, 2010.

[2] Dana Berman, Shai Avidan, et al. Non-local image dehazing. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1674–1682, 2016.

[3] Simone Bianco, Gianluigi Ciocca, and Raimondo Schettini. How far can you get by combining change detection algorithms? In *International Conference on Image Analysis and Processing*, pages 96–107. Springer, 2017.

[4] Thierry Bouwmans and El Hadi Zahzah. Robust pca via principal component pursuit: A review for a comparative evaluation in video surveillance. *Computer Vision and Image Understanding*, 122:22–34, 2014.

[5] Yuri Boykov and Vladimir Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE transactions on pattern analysis and machine intelligence*, 26(9):1124–1137, 2004.

[6] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy

minimization via graph cuts. *IEEE Transactions on pattern analysis and machine intelligence*, 23(11):1222–1239, 2001.

[7] Ricardo Cabral, Fernando De la Torre, João P Costeira, and Alexandre Bernardino. Unifying nuclear norm and bilinear factorization approaches for low-rank matrix decomposition. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2488–2495, 2013.

[8] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198, 2016.

[9] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011.

[10] Laurent Caraffa and Jean-Philippe Tarel. Stereo reconstruction and contrast restoration in daytime fog. In *Asian conference on computer vision*, pages 13–25. Springer, 2012.

[11] Deniz Engin, Anil Genç, and Hazim Kemal Ekenel. Cycle-dehaze: Enhanced cyclegan for single image dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 825–833, 2018.

[12] Pedro F Felzenszwalb and Daniel P Huttenlocher. Efficient graph-based image segmentation. *International journal of computer vision*, 59(2): 167–181, 2004.

[13] Guy Gilboa and Stanley Osher. Nonlocal operators with applications to image processing. *Multiscale Modeling & Simulation*, 7(3):1005–1028, 2009.

BIBLIOGRAPHY

[14] Alona Golts, Daniel Freedman, and Michael Elad. Unsupervised single image dehazing using dark channel prior loss. *IEEE Transactions on Image Processing*, 2019.

[15] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. Weighted nuclear norm minimization with application to image denoising. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2862–2869, 2014.

[16] Kaiming He, Jian Sun, and Xiaoou Tang. Guided image filtering. In *European conference on computer vision*, pages 1–14. Springer, 2010.

[17] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353, 2010.

[18] Wei He, Yongkwan Kim, Jianhui Wu, Guoyun Zhang, Qi Qi, Longyuan Guo, Bing Tu, and Feng Huang. Local compact binary patterns for background subtraction in complex scenes. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 1518–1523. IEEE, 2018.

[19] Marko Heikkila and Matti Pietikainen. A texture-based method for modeling the background and detecting moving objects. *IEEE transactions on pattern analysis and machine intelligence*, 28(4):657–662, 2006.

[20] Martin Hofmann, Philipp Tiefenbacher, and Gerhard Rigoll. Background segmentation with feedback: The pixel-based adaptive segmenter. In *2012 IEEE computer society conference on computer vision and pattern recognition workshops*, pages 38–43. IEEE, 2012.

[21] Shih-Chia Huang, Bo-Hao Chen, and Wei-Jheng Wang. Visibility restoration of single hazy images captured in real-world weather con-

ditions. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(10):1814–1824, 2014.

[22] Shih-Shinh Huang, Li-Chen Fu, and Pei-Yung Hsiao. Region-level motion-based background modeling and subtraction using mrfs. *IEEE Transactions on Image Processing*, 16(5):1446–1456, 2007.

[23] Sajid Javed, Seon Ho Oh, Andrews Sobral, Thierry Bouwmans, and Soon Ki Jung. Or-pca with mrf for robust foreground detection in highly dynamic backgrounds. In *Asian Conference on Computer Vision*, pages 284–299. Springer, 2014.

[24] Eunwoo Kim, Minsik Lee, and Songhwai Oh. Elastic-net regularization of singular values for robust subspace learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 915–923, 2015.

[25] Harald Koschmieder. Theorie der horizontalen sichtweite. *Beitrage zur Physik der freien Atmosphare*, pages 33–53, 1924.

[26] Anat Levin, Dani Lischinski, and Yair Weiss. A closed-form solution to natural image matting. *IEEE transactions on pattern analysis and machine intelligence*, 30(2):228–242, 2007.

[27] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4770–4778, 2017.

[28] Liyuan Li, Weimin Huang, Irene YH Gu, and Qi Tian. Foreground object detection from videos containing complex background. In *Proceedings of the eleventh ACM international conference on Multimedia*, pages 2–10, 2003.

[29] Runde Li, Jinshan Pan, Zechao Li, and Jinhui Tang. Single image dehazing via conditional generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8202–8211, 2018.

[30] Yi Li. Calculation of smoke transmission based on first-order multiple scattering approximation. *CHINESE JOURNAL OF EXPLOSIVES AND PROPELLANTS*, 24(2):39–41, 2001.

[31] Long Ang Lim and Hacer Yalim Keles. Learning multi-scale features for foreground segmentation. *Pattern Analysis and Applications*, pages 1–12, 2019.

[32] Zhouchen Lin, Minming Chen, and Yi Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*, 2010.

[33] Qi Liu, Xinbo Gao, Lihuo He, and Wen Lu. Single image dehazing with depth-aware non-local total variation regularization. *IEEE Transactions on Image Processing*, 27(10):5178–5191, 2018.

[34] Ramavtar Malav, Ayoung Kim, Soumya Ranjan Sahoo, and Gaurav Pandey. Dhsgan: An end to end dehazing network for fog and smoke. In *Asian Conference on Computer Vision*, pages 593–608. Springer, 2018.

[35] Murari Mandal, Prafulla Saxena, Santosh Kumar Vipparthi, and Subrahmanyam Murala. CANDID: Robust change dynamics and deterministic update policy for dynamic background subtraction. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 2468–2473. IEEE, 2018.

[36] Gaofeng Meng, Ying Wang, Jiangyong Duan, Shiming Xiang, and Chunhong Pan. Efficient image dehazing with boundary constraint and

contextual regularization. In *Proceedings of the IEEE international conference on computer vision*, pages 617–624, 2013.

[37] Hossein Ziaei Nafchi and Mohamed Cheriet. Efficient no-reference quality assessment and classification model for contrast distorted images. *IEEE Transactions on Broadcasting*, 64(2):518–523, 2018.

[38] Srinivasa G Narasimhan and Shree K Nayar. Vision and the atmosphere. *International journal of computer vision*, 48(3):233–254, 2002.

[39] John P Oakley and Brenda L Satherley. Improving image quality in poor visibility conditions using a physical model for contrast degradation. *IEEE transactions on image processing*, 7(2):167–179, 1998.

[40] Tae-Hyun Oh, Yu-Wing Tai, Jean-Charles Bazin, Hyeongwoo Kim, and In So Kweon. Partial sum minimization of singular values in robust pca: Algorithm and applications. *IEEE transactions on pattern analysis and machine intelligence*, 38(4):744–758, 2015.

[41] Timo Ojala, Matti Pietikäinen, and Topi Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (7):971–987, 2002.

[42] Chong Peng, Chenglizhao Chen, Zhao Kang, Jianbo Li, and Qiang Cheng. Res-pca: A scalable approach to recovering low-rank matrices. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7317–7325, 2019.

[43] John H Reif. Minimum s-t cut of a planar undirected network in o(n\logˆ2(n)) time. *SIAM Journal on Computing*, 12(1):71–81, 1983.

[44] Wenqi Ren, Lin Ma, Jiawei Zhang, Jinshan Pan, Xiaochun Cao, Wei Liu, and Ming-Hsuan Yang. Gated fusion network for single image de-

hazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3253–3261, 2018.

[45] Jesus Dario Romero, María J Lado, and Arturo J Méndez. A background modeling and foreground detection algorithm using scaling coefficients defined with a color model called lightness-red-green-blue. *IEEE Transactions on Image Processing*, 27(3):1243–1258, 2017.

[46] Saad Bin Sami, Abdul Muqeet, and Humera Tariq. A novel image dehazing and assessment method. *ArXiv*, abs/2001.06963, 2019.

[47] MS Sandeep. Remote sensing image dehazing using guided filter. *IJRSCSE*, 1(3):44–49, 2014.

[48] Ja-Won Seo and Seong Dae Kim. Dynamic background subtraction via sparse representation of dynamic textures in a low-dimensional subspace. *Signal, Image and Video Processing*, 10(1):29–36, 2016.

[49] Fanhua Shang, James Cheng, Yuanyuan Liu, Zhi-Quan Luo, and Zhouchen Lin. Bilinear factor matrix norm minimization for robust pca: Algorithms and applications. *IEEE transactions on pattern analysis and machine intelligence*, 40(9):2066–2080, 2017.

[50] Yuan Shen, Zaiwen Wen, and Yin Zhang. Augmented lagrangian alternating direction method for matrix separation based on low-rank factorization. *Optimization Methods and Software*, 29(2):239–263, 2014.

[51] Bijan Shoushtarian and Helmut E Bez. A practical adaptive approach for dynamic background subtraction using an invariant colour model and object tracking. *Pattern Recognition Letters*, 26(1):5–26, 2005.

[52] Pierre-Luc St-Charles, Guillaume-Alexandre Bilodeau, and Robert Bergevin. SuBSENSE: A universal change detection method with lo-

cal adaptive sensitivity. *IEEE Transactions on Image Processing*, 24(1): 359–373, 2014.

[53] Chris Stauffer and W Eric L Grimson. Adaptive background mixture models for real-time tracking. In *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, volume 2, pages 246–252. IEEE, 1999.

[54] Zhiming Tan, Xianghui Bai, Bingrong Wang, and Akihiro Higashi. Fast single-image defogging. *Fujitsu Sci. Tech. J*, 50(1):60–65, 2014.

[55] Kentaro Toyama, John Krumm, Barry Brumitt, and Brian Meyers. Wallflower: Principles and practice of background maintenance. In *Proceedings of the seventh IEEE international conference on computer vision*, volume 1, pages 255–261. IEEE, 1999.

[56] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9446–9454, 2018.

[57] Kunfeng Wang, Chao Gou, and Fei-Yue Wang. M^4CD: A robust change detection method for intelligent visual surveillance. *IEEE Access*, 6: 15505–15520, 2018.

[58] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Transactions on Image Processing*, 22(9):3538–3548, 2013.

[59] Yi Wang, Pierre-Marc Jodoin, Fatih Porikli, Janusz Konrad, Yannick Benezeth, and Prakash Ishwar. Cdnet 2014: an expanded change detection benchmark dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 387–394, 2014.

[60] Yuanyu Wang, Yuanzong Li, and Tianxu Zhang. Study on the method of image restoration in the environment of dust. In *2010 6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM)*, pages 1–4. IEEE, 2010.

[61] Zhunxuan Wang, Zipei Wang, Qiqi Li, and Hakan Bilen. Image deconvolution with deep image and kernel priors. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 0–0, 2019.

[62] Christopher Richard Wren, Ali Azarbayejani, Trevor Darrell, and Alex Paul Pentland. Pfinder: Real-time tracking of the human body. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7): 780–785, 1997.

[63] Jia Yan, Jie Li, and Xin Fu. No-reference quality assessment of contrast-distorted images using contrast enhancement. *arXiv preprint arXiv:1904.08879*, 2019.

[64] Miao Yang and Arcot Sowmya. An underwater color image quality evaluation metric. *IEEE Transactions on Image Processing*, 24(12):6062–6071, 2015.

[65] Xiaoming Yuan and Junfeng Yang. Sparse and low-rank matrix decomposition via alternating direction methods. *preprint*, 12(2), 2009.

[66] He Zhang and Vishal M Patel. Densely connected pyramid dehazing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3194–3203, 2018.

[67] Shengping Zhang, Hongxun Yao, and Shaohui Liu. Dynamic background modeling and subtraction using spatio-temporal local binary patterns. In *2008 15th IEEE International Conference on Image Processing*, pages 1556–1559. IEEE, 2008.

[68] Tao Zhang, Changyan Shao, and Xinnian Wang. Atmospheric scattering-based multiple images fog removal. In *2011 4th International Congress on Image and Signal Processing*, volume 1, pages 108–112. IEEE, 2011.

[69] Xiaowei Zhou, Can Yang, and Weichuan Yu. Moving object detection by detecting contiguous outliers in the low-rank representation. *IEEE transactions on pattern analysis and machine intelligence*, 35(3):597–610, 2012.

[70] Mingzhu Zhu, Bingwei He, and Qiang Wu. Single image dehazing based on dark channel prior and energy minimization. *IEEE Signal Processing Letters*, 25(2):174–178, 2017.

[71] Zoran Zivkovic. Improved adaptive gaussian mixture model for background subtraction. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, volume 2, pages 28–31. IEEE, 2004.

국문초록

강건 주성분 분석은 배경 감산을 통한 동영상의 전경 추출의 방법으로 이용되어 왔으나, 동적 배경은 저계수 행렬로 표현될 수 없기 때문에 동적 배경 감산에 성능적 한계를 가지고 있었다. 우리는 전경과 배경을 구분하는 일계 마르코프 연쇄를 도입해 정적 배경을 나타내는 항과 곱하고 이것을 이용한 새로운 형태의 강건 주성분 분석을 제안하여 동적 배경 감산 문제를 해결한다. 해당 최소화 문제는 반복적인 교차 최적화를 통하여 해결한다. 이어서 대기중의 미세먼지에 의해 오염된 영상을 복원한다. 영상 분할과 암흑 채널 가정에 기반하여 깊이 지도를 구하고, 비국소 총 변동 최소화를 통하여 정제한다. 이후 깊은 영상 가정에 기반한 영상 생성기를 통하여 최종적으로 복원된 영상을 구한다. 실험을 통하여 제안된 방법을 다른 방법들과 비교하고 질적인 측면과 양적인 측면 모두에서 우수함을 확인한다.

주요어휘: 이동 물체 탐지, 동적 배경 감산, 강건한 주성분 분석, 영상 먼지 제거, 깊은 영상 가정
학번: 2015-20267

감사의 글

하나부터 열까지 사려깊게 챙겨주신 강명주 지도교수님 감사합니다. 덕분에 제가 다양한 경험을 하며 발전할 수 있었습니다. 금융수학을 공부하던 학부생 때부터 많은 기회 접하게 해주신 이기암 교수님, 대학원에서 영상처리의 기초를 닦아주신 강명민 교수님 역시 감사합니다. 귀한 시간 내주셔서 논문심사 맡아주신 국웅 교수님, 정미연 교수님께도 감사합니다.

같이 연구하고 일했던 NCIA 구성원분들께도 감사합니다. 특히 영상처리 같이 공부했던 선배님들, 같이 연구하고 프로젝트 해왔던 동기들과 후배님들 모두 좋은 성과 내길 바랍니다.

마지막으로 언제나 절 믿고 지지해준 아버지, 어머니, 두 누님께 감사의 말씀을 드리고 싶습니다.