Journal of Korea Multimedia Society Vol. 22, No. 10, October 2019(pp. 1149-1159)
https://doi.org/10.9717/kmms.2019.22.10.1149

단일 영상 비균일 블러 제거를 위한 다중 학습 구조

정형주[†], 장현성[††], 하남구[†††], 연윤모[††††], 권구용[††††], 손광훈[††††††]

Multi-task Architecture for Singe Image Dynamic Blur Restoration and Motion Estimation

Hyungjoo Jung[†], Hyunsung Jang[††], Namkoo Ha[†††], Yoonmo Yeon[††††],
Ku yong Kwon[†††††], Kwanghoon Sohn[††††††]

ABSTRACT

We present a novel deep learning architecture for obtaining a latent image from a single blurry image, which contains dynamic motion blurs through object/camera movements. The proposed architecture consists of two sub-modules: blur image restoration and optical flow estimation. The tasks are highly related in that object/camera movements make cause blurry artifacts, whereas they are estimated through optical flow. The ablation study demonstrates that training multi-task architecture simultaneously improves both tasks compared to handling them separately. Objective and subjective evaluations show that our method outperforms the state-of-the-arts deep learning based techniques.

Key words: Dynamic Motion Deblurring, CNN, Motion Estimation, Multi-task Architecture

1. 서 론

일반적으로 사진은 특정 시간에 즉각적인 기억을 포착한다고 생각되지만, 이는 기술적으로 사실이 아니다. 카메라의 센서는 장면에 대한 정보를 얻기 위해 빛을 축적하며, 이때 충분한 노출 시간이 필요하다. 그 결과, 노출 시간 중의 카메라/물체의 움직임은 사진에서 모션 블러를 발생시킨다. 카메라/물체의 움직임에 의한 모션 블러는 사진의 가시성을 떨어뜨리며, 이는 다양한 물체 인식 [1-3] 분야의 성능을 저하한다. 이러한 모션 블러에 의해 열화된 영상을 복원하는 것은 가시적으로 향상된 영상을 얻을 수 있으며, 다양한 객체 탐지 분야에서 매우 필수적인 전처리 과정이다.

기존의 연구들 [4-7]은 다음의 블러 영상 생성 모델을 기반으로 연구가 진행되었다.

$$B = KS + n, \tag{1}$$

식 (1)에서 B, S, 그리고 n은 각각 벡터화된 블러 영상, 뚜렷한 영상, 그리고 additive 노이즈를 의미한다. K는 블러 커널을 나타내는 매우 큰 크기의 드문

※ Corresponding Author : Kwanghoon Sohn, Address: (120-749) Digital Image Media Labatory. C129, The 3rd Engineering Building, Yonsei University 50 Yonsei-ro, Seodaemun-Gu, Seoul 120-749, Korea, TEL : +82-2-2123-2879, FAX : +82-2-2123-7712, E-mail : khsohn@yonsei.ac.kr
Receipt date : Aug. 12, 2019, Revision date : Sep. 17, 2019
Approval date : Sep. 26, 2019
[†] School of Electrical Electronic Engineering, Yonsei University (E-mail : coolguy0220@yonsei.ac.kr)
[††] EO/IR R&D Lab., LIG Nex1 Co., Ltd., 16911, Yongin-city, South Korea (E-mail : hyunsung.jang@gmail.com)
[†††] EO/IR R&D Lab., LIG Nex1 Co., Ltd., 16911, Yongin-city, South Korea (E-mail : hanamkoo369a@lignex1.com)
[††††] EO/IR R&D Lab., LIG Nex1 Co., Ltd., 16911, Yongin-city, South Korea (E-mail : yoonmo.yeon@lignex1.com)
[†††††] EO/IR R&D Lab., LIG Nex1 Co., Ltd., 16911, Yongin-city, South Korea (E-mail : kuyon.kwon@lignex1.com)
[††††††] School of Electrical Electronic Engineering, Yonsei University

행렬 (Sparse matrix)을 나타낸다. 일반적으로 블러 커널 K는 알려진 값이 아니며, 기존의 블라인드 디컨볼루션 (Blind deconvolution) 방식은 K와 S를 동시에 추정하여 영상 개선 영상을 획득한다. 식 (1)의 모델을 기반으로 영상의 모든 픽셀에 대한 블러 커널을 찾는 것은 ill-posed 문제이다. 기존의 방식 [4-7]들은 이러한 ill-posed 문제를 해결하기 위하여 블러 커널에 다양한 가정을 설정하였다. 기존의 다양한 연구는 블러 커널 K를 균일한 블러 커널로 가정하였다. 그러나 균일한 블러 커널은 카메라의 회전 변화 및 객체의 움직임을 근사화하기에 부정확한 문제가 있다. 이러한 문제를 해결하기 위하여 비균일 블러 커널을 추정하기 위한 연구 [8, 9]가 많이 진행되었다. 그 결과 드문 행렬 K가 밀도가 매우 커졌으며, 이는 영상 개선을 위해 매우 많은 연산량이 필요하다. 또한, 기존의 비균일 블러 제거 기술은 객체의 움직임에 의한 동적인 블러 문제에 잘 적용하지 못한다.

최근 다양한 물체 인식 분야 [1-3]에서 딥 러닝 기술의 발전이 후, 컨볼루션 신경망 (Convolutional neural networks, CNNs)을 이용한 다양한 영상 개선 연구 (예: 노이즈 제거 [10], 블러 제거 [11], 초고해상도 복원 [12], 및 비 제거 [13] 등) 가 진행되고 있다. CNNs을 영상 블러 제거에 적용하기 위해서는 블러 영상과 이에 대응하는 뚜렷한 영상 쌍을 구축하는 것이 가장 필수적이다. 초기의 연구 [11]에서는 공간적으로 불변 (invariant) 또는 변하는 블러 커널을 이용하여 뚜렷한 영상을 열화하여 블러 영상을 생성하였다. 그러나 이러한 인위적 영상은 객체의 움직임에 의한 동적인 블러를 자연스럽게 시뮬레이션할 수 없다. 최근 블러 커널을 이용한 영상 생성 대신, GoPro 카메라를 이용한 자연스러운 비균일 블러 영상을 포함하는 데이터베이스 [14]가 구축되었다. 최신의 CNNs 기반의 연구는 이를 이용하여 입력 블러 영상과 뚜렷한 정답 영상 간의 높은 차원의 비선형 매핑 관계를 학습한다. 이를 위해 멀티스케일 구조 [14] 및 Generative adversarial network (GAN) [15] 등의 구조가 많이 사용된다.

본 논문에서는 영상 획득 시 동적 블러가 생성되는 원인을 고려해, 객체의 움직임에 의한 단일 비균일 블러 영상을 개선하기 위한 다중 학습 구조를 제안한다. 제안하는 구조는 단일 블러 영상으로부터 영상 개선을 수행할 뿐만 아니라, 모션 정보를 옵티컬 플로우의 형태로 추정한다. 첫 번째로 블러의 생성 원인이 객체/카메라의 모션에 의한 것이므로 모션 추정은 영상 개선에 도움이 된다. 한편 단일 블러 영상에서의 모션 추정을 위한 단서는 블러의 정도이다. 블러가 심한 영역은 큰 모션 정보를 갖으며, 반대로 뚜렷한 영역은 적은 모션을 포함한다. 각각의 테스크는 네트워크의 파라미터 공유를 통해 상호보완적으로 서로의 성능을 향상한다. 본 논문은 다양한 실험을 통해 제안하는 블러 제거 및 모션 추정의 다중 학습 구조가 단일 블러 제거 만을 수행하였을 때보다 좋은 성능을 보임을 확인한다. 또한, 최신의 state-of-the-arts 방식과의 비교를 통해 제안하는 방식의 객관적/주관적 성능을 평가한다.

본 논문의 구성은 다음과 같다. 2장에서 본 논문과 관련된 블러 개선 기술과 옵티컬 플로우 추정 기술을 설명하며, 3장에서는 자연스러운 동적 블러 영상 생성 방법을 소개한다. 그리고 4장에서는 제안하는 알고리듬의 네트워크 구조 및 비용 함수를 서술한다. 5장에서는 제안하는 기술의 분석과 최신 기술과의 객관적/주관적 비교하여 성능을 평가하며, 6장에서 결론을 맺는다.

2. 관련 연구

2.1 기존의 커널 추정 기반 블러 제거 기술

커널 추정 기반의 블러 제거 기술은 주로 최적화 문제 기반으로 연구가 되었다. 좋은 성능을 블러 영상 개선을 위해 최적화 기반의 기술은 뚜렷한 영상의 구조적 정보를 나타낼 수 있는 가정을 필요로 한다. 대표적인 영상 구조에 대한 가정은 그래디언트의 드문 정도 [4], hyper-Laplacian 가정 [5], L0-norm 가정 [6], normalized 드문 정도 [7] 등이 있다. 그러나 영상 구조에 대한 이러한 휴리스틱한 가정은 기존의 최적화 기반 블러 제거 기술이 일반적인 영상에 대해 좋은 성능을 보이지 못하게 한다. 기존의 커널 추정 기반의 기술은 영상의 구조에 대한 가정뿐만 아니라, 블러 커널에 대한 가정을 포함한다. 입력 블러 영상으로부터 블러 커널을 추정하는 것은 매우 ill-posed 문제이며, 이러한 어려움을 보완하기 위해 블러 커널에 대한 다양한 가정을 이용한다. 대표적인 가정으로 영상 전체에 대하여 공간적으로 균일한 블러 커널을 가정한다. 그러나 이는 객체의 움직임에 의한 동적인

블러에 적합하지 않다. 이를 해결하기 위하여, Kim 등[8]은 영상 분할과 블러 제거를 동시에 수행하는 동적 블러 영상 개선 기술을 제안하였다. 공간적으로 불균일한 블러 문제를 해결하기 위해 영상의 분할 영역에서 비선형의 블러 커널을 각각 추정하였다. 또한, Kim 등[9]은 지역적으로 선형의 블러 커널을 추정하였으며, 뚜렷한 영상과 선형적 모션 정보를 동시에 추정하였다. 그러나 이러한 가정은 일반적인 영상에서 정확한 블러 커널의 추정을 어렵게 하였으며, 그 결과 동적 블러 영역을 갖는 영상에서 좋은 성능을 획득할 수 없으며 많은 ringing 결함이 발생한다.

2.2 CNNs 기반의 블러 제거 기술

최근 다양한 CNNs 기반의 블러 제거 기술이 제안되었다. CNNs 기반의 블러 제거 기술은 크게 커널 추정 방식과 커널 추정 없이 바로 개선 영상을 생성하는 방식으로 구분된다. 커널 추정 방식의 대표적인 예로, Xu 등[11]은 CNNs을 이용하여 주어진 블러 커널 기반의 논블라인드 블러 제거 기술을 제안하였다. 그들은 블러 커널을 몇 개의 작은 커널을 분리할 수 있는 커널 분리 특성을 고려한 네트워크를 설계하였다. 또한, 네트워크는 노이즈나 컬러 포화 문제를 줄일 수 있는 디노이징 기능을 수행한다. 한편, Schuler 등[16]은 기존의 커널 추정 기반의 블라인드 블러 제거 기술을 CNNs을 이용하여 구현하였다. 그들은 표현자 추출, 커널 추정 및 논블라인드 블러 제거 기술을 반복적으로 CNNs을 사용하였다. 이러한 초기의 CNNs 기반 블러 제거 기술 [11, 16]은 네트워크를 학습하기 위해 인공의 균일한 블러 커널을 사용하여 입력 블러 영상을 생성하였다. 그러나 이러한 기술은 객체의 움직임에 의한 동적 블러를 제거하는데 부적합하다. 비균일 블러 제거를 위해, Sun 등[17]은 73개의 후보 블러 커널을 혼합하여 영상의 부분적 패치에 서로 다른 블러를 생성하였다. 네트워크는 73개의 후보 커널 중 어떠한 커널이 작용하였는지를 분류한다. 그러나 커널 추정 기반 블러 제거 기술은 몇 가지 문제점을 갖고 있다. 첫째로 간단한 커널의 가정은 폐쇄된 영역이나 깊이 변화와 같은 몇 가지 어려운 상황을 정확히 모델링 할 수 없다. 둘째로 커널 추정 과정의 블러 커널 및 뚜렷한 개선 영상에 대한 휴리스틱한 가정이 필요하다. 마지막으로 객체의 동적 블러를 포함하는 영상에서 모든 픽셀에 대해 공간적으로 변화하는 커널을 찾으려면 많은 양의 메모리와 계산이 필요하다.

최근 커널 추정 없이 CNNs을 이용하여 동적 블러 영상을 개선하는 기술이 많이 연구되었다. 강력한 end-to-end 학습과 높은 차원의 비선형 구조는 CNNs이 블러 제거에 있어 좋은 성능을 보이게 한다. Nah 등[14]은 블러 영상을 개선하기 위해 세 스케일로 구성된 CNNs 구조를 제안하였다. 제안하는 구조는 작은 스케일에서 큰 스케일로 블러된 영상을 개선한다. DeblurGAN [15]은 블러된 영상으로부터 영상을 개선하기 위한 GAN [18] 구조를 제안하였다. Tao 등[19]은 long short-term memory (LSTM)의 회귀 구조를 이용한 멀티스케일 블러 제거 구조를 제안하였다. Zhang 등[20]은 공간적으로 변하는 recurrent neural networks (RNNs)을 이용한 블러 제거 네트워크 구조를 제안하였다.

2.3 비디오 영상의 움직임 추정 기술

옵티컬 플로우 (Optical flow)의 형태로 움직임 정보를 추정하는 것은 다양한 비디오 응용 분야에 있어서 핵심적인 기술이다. 기존의 연구는 최적화 문제를 통한 옵티컬 플로우 추정 연구가 많이 이루어졌다. Brox 등[21]은 Euler-Lagrange 알고리즘을 통해 최적화 문제를 해결하였다. 또한, 큰 움직임 변화를 추정하기 위하여 다양한 표현자 기반의 매칭 기술이 최적화 문제와 결합하였다. 예를 들어 SIFT-flow [22]와 PatchMatch [23] 기술의 경우 작은 패치의 크기에서 정합을 수행한다. Xu 등[24]은 SIFT와 Patch Match를 통합하여 좋은 성능을 얻었다. EpicFlow [25]는 드문 매칭 (Sparse matching)된 영역을 보간하여 모든 픽셀에서의 초기 정합을 추정하며, 이때 경계보존 보간을 통해 추정된 옵티컬 플로우와 입력 영상의 경계선을 일치시켰다. 그 후, 보간된 초기 정합을 이용하여 기존의 최적화 문제를 설계하여 모든 픽셀에서의 정합을 수행한다. 그러나 기존의 최적화 문제를 통한 옵티컬 플로우 추정 기술은 많은 연산량을 필요한 문제점이 있다.

최근 CNNs을 이용한 옵티컬 플로우 추정 연구가 많이 제안되었다. CNNs의 강력한 표현자 추출 능력을 통해 최고의 성능을 보인다. 학습 기반 옵티컬 플로우 추정의 한 방식은 영상 패치에서 정합을 수행하기 위해 CNNs을 사용하는 것이다. Thewlis 등[26]

은 Deep Matching [27]을 보완하여 end-to-end 학습이 가능한 구조를 제안하였다. 또한, Gadot 등[28]은 Siamese 네트워크 구조를 사용하여 영상 패치에서 표현자를 학습한다. 이러한 방식은 정확도는 높지만, 많은 연산량이 필요해 실제 응용 기술에 적용되기 어렵다. 입력 영상 쌍으로부터 정합 결과를 바로 추정하는 FlowNet [29]은 대표적인 CNNs 기반 옵티컬 플로우 추정 기술의 하나로, CNNs을 이용한 end-to-end 학습을 통해 네트워크를 학습한다. FlowNet 이후 CNNs 기반의 옵티컬 플로우 추정의 다양한 연구가 진행되었다. 이는 3D 콘볼루션 네트워크 [30], 비지도학습 기술 [31], 회전 변화에 강인한 네트워크 구조 [32], 그리고 피라미드 구조 [33] 등의 연구가 수행되었다.

3. 동적 블러 영상 생성

CNNs을 이용하여 객체/카메라의 움직임에 의한 동적 블러 영상을 개선하기 위해서 가장 중요한 것은 자연스러운 블러 영상을 생성하는 것이다. 이를 위해 최근 카메라 센서의 영상 획득 방식을 고려한 데이터베이스 [14]가 구축되었다. 카메라 센서는 노출 시간 동안 장면에 대한 정보를 획득하기 위하여 빛 정보를 촬영하며, 노출 시간 동안의 객체/카메라의 움직임에 의해 블러가 발생한다. 카메라를 통해 획득되는 정보는 그 후 각 픽셀 단위로 비선형적 카메라 반응 함수 (Camera response function, CRF)에 의해 변형된다. 그 결과 카메라의 영상 획득 모델은 다음 식 (2)로 모델링 된다.

$$B = g\left(\frac{1}{T}\int_0^T S(t)dt\right), \tag{2}$$

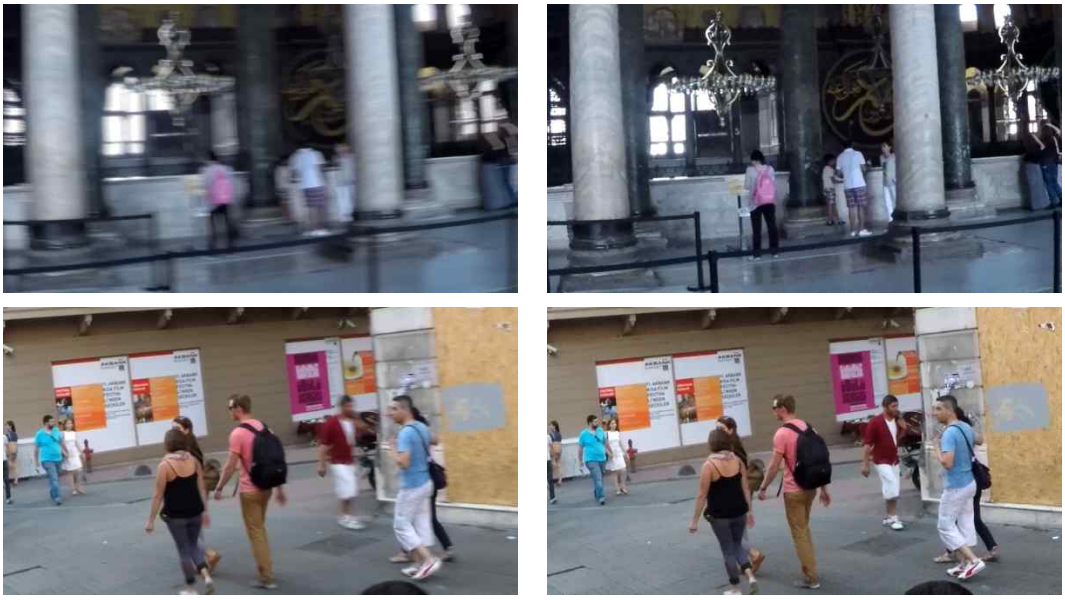
이때, T와 $S(t)$는 각각 노출 시간과 t 시각에 획득된 센서의 영상 신호를 나타내며, $g(\bullet)$는 카메라의 CRF를 의미한다. [14]의 데이터베이스는 식 (2)의 영상 획득 모델을 이산 시간으로 근사하기 위하여 다음과 같이 블러 영상을 모델링 하였다.

$$B = g\left(\frac{1}{M}\sum_{i=0}^{M-1} S[i]\right), \tag{3}$$

식 (3)에서 M과 $S[i]$는 노출 시간 동안의 영상 샘플 수 및 i번째 영상 샘플을 의미한다. 일반적으로 비선형 CRF에 대한 정답을 추정하기 매우 어려우며, [14] 데이터베이스에서는 다음 식 (4)의 감마 조절을 통해 블러 영상을 획득하였다($\gamma = 2.2$).

$$g(x) = x^{1/\gamma}, \tag{4}$$

[14]의 데이터베이스는 GoPro4 Hero Black 카메



(a) (b)

Fig. 1. (a) Blurry images generated by (3), (b) Ground truth sharp images.

라를 이용하여 영상을 획득하였다. GoPro 카메라를 이용하여 240fps의 동영상을 촬영하였으며, 연속적인 뚜렷한 프레임 (7−13)을 평균하여 객체의 동적인 움직임에 의한 블러를 포함하는 자연스러운 영상을 생성하였다. Fig. 1은 본 논문에서 네트워크 학습 및 성능 평가에 사용한 [14] 데이터베이스의 샘플 영상을 보여준다. Fig. 1-(a)는 7장의 뚜렷한 영상을 평균하여 생성된 동적 블러를 포함하는 영상이고, Fig. 1-(b)는 중간 프레임의 뚜렷한 영상이다.

4. 제안한 방법

4.1 제안한 알고리듬의 개요

제안하는 다중 학습 구조는 단일 블러 영상을 입력으로 하여, 각각 블러가 제거된 영상과 옵티컬 플로우를 동시에 추정한다. 네트워크 파라미터 공유를 통하여 두 테스크의 성능을 향상할 수 있다. 먼저 옵티컬 플로우 추정 네트워크는 입력 영상으로부터 모션 정보를 추정하기 때문에, 블러 영상 개선 네트워크의 성능을 향상한다. 또한, 블러 개선 네트워크를 통해 학습된 표현자는 단일 영상에서 옵티컬 플로우를 추정할 때, 블러의 정도를 분별하는 데 도움을 줄 수 있다. Fig. 2는 제안하는 전체적인 다중 학습 구조의 흐름도를 나타낸다.

4.2 다중 학습 네트워크 구조

제안하는 다중 학습 구조는 Fig. 2와 같이 크게 블러 영상 개선 네트워크와 옵티컬 플로우 추정 네트워크로 구성된다. 두 네트워크 모두 넓은 범위의 영역을 참조하기 위해 strided 콘볼루션을 통한 encoder-decoder 구조로 설계하였으며, 서로 일부 네트워크 파라미터를 공유한다. 블러 개선 네트워크는 먼저 단일 입력 영상으로부터 세 개의 콘볼루션 레이어를 통해 64 채널의 표현자를 추출한다. 그 후, strided 콘볼루션을 포함하는 두 개의 ResBlock [1]을 통해 공간적 크기를 반으로 줄이면서 동시에 표현자 채널의 수를 2배씩 증가시킨다. ResBlock은 네트워크 학습 시 그래디언트 사라지는 문제를 해결할 수 있어 더욱 좋은 표현자를 학습할 수 있는 장점이 있다. ResBlock은 세 개의 콘볼루션 레이어로 구성되며, 첫 번째 레이어는 stride가 2이다. 또한, ResBlock의 결과는 마지막 콘볼루션 레이어의 결과와 첫 번째 레이어의 결과를 합하여 구한다. 1/4 크기로 줄어든 표현자 (채널은 256)는 그 후 세 개의 ResBlock을 통과하며, 이때 모든 콘볼루션 레이어의 stride는 1이다. 마지막으로 입력 영상과 동일한 해상도를 갖는 개선 영상을 추정하기 위해 두 개의 디콘볼루션 [34] 레이어를 사용하여 해상도를 2배씩 증가시킨다. 각 디콘볼루션 레이어 이후 한 개의 콘볼루션 레이어를 추가로 설계하였으며, 표현자의 채널 수는 encoder와 대칭이 되도록 반으로 감소시켰다. 마지막 콘볼루션 레이어는 세 채널의 개선 영상을 바로 추정한다. 한편 대칭이 되는 encoder와 decoder 표현자 간 skip 연결을 통해 앞 단의 표현자의 공간적 구조를 유지 시켰다. 네트워크의 비선형성을 높이기 위한 비선형 함수로 Rectified linear unit (ReLU) 레이어를 사용하였다. 또한, 콘볼루션 표현자의 내부 코베리언스 변화 (Internal covariate shift) 문제를 해

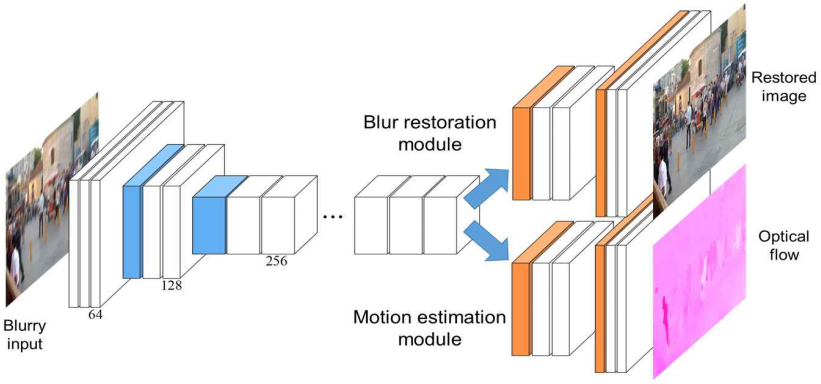


Fig. 2. The overall framework of the proposed multi−task architecture for simultaneous deblurring and optical flow estimation.

결하기 위한 Batch normalization [35] 레이어를 사용하였다.

다중 학습 구조의 다른 네트워크인 옵티컬 플로우 추정 네트워크는 블러 영상 개선 네트워크의 encoder 구조를 공유한다. 옵티컬 플로우 추정을 위한 추가적인 네트워크는 decoder 부분의 디콘볼루션 레이어와 콘볼루션 레이어이다. 이는 블러 개선 네트워크의 decoder 부분과 유사하게 설계하였다. 옵티컬 플로우 추정 네트워크의 마지막 콘볼루션은 네 채널의 옵티컬 플로우를 바로 추정한다. 각각의 두 채널 짜리 플로우는 각각 forward와 backward 플로우를 나타낸다.

4.3 비용 함수 (Loss function)

제안하는 다중 학습 구조를 학습하기 위한 전체 비용 함수 L_t는 블러 영상 개선 및 옵티컬 플로우 추정 네트워크를 학습하기 위한 L_{deblur}와 L_{flow}의 비용 함수의 합으로 설계된다. L_{deblur}는 최신의 CNNs 기반 초고해상도 영상 복원 기술 [10-13]을 따라 다음의 perceptual loss와 adversarial loss의 합으로 설계하였다 ($\lambda = 0.01$).

$$L_{deblur} = L_{VGG} + \lambda L_{adv},. \tag{5}$$

L_{VGG}는 ImageNet 데이터베이스를 이용한 영상 분류 분야에 이미 학습된 VGG 네트워크 [36]의 표현자를 통해 계산된다. 본 논문에서는 Conv3_3의 표현자를 사용하였으며, L_{VGG}는 개선 영상 S_{deblur}와 정답 영상 $S_{\lfloor M/2 \rfloor}$를 이용하여 다음과 같이 설계된다.

$$L_{VGG} = \| \Phi(S_{deblur}) - \Phi(S_{\lfloor M/2 \rfloor}) \|^2, \tag{6}$$

이때, Φ는 VGG 네트워크의 Conv3_3 표현자 추정 네트워크를 나타낸다. 이러한 의미론적 특성을 갖는 표현자를 이용한 비용 함수는 가시적으로 향상된 개선 영상을 획득할 수 있다. L_{adv}는 개선 영상과 정답 영상 간의 구분이 안 되도록 하는 비용 함수이다. 이러한 adversarial 학습을 위해서는 입력 영상이 실제 뚜렷한 영상인지 개선된 영상인지를 분류하기 위한 discriminator 네트워크가 필요하다. 본 논문에서는 기존의 GAN 기술을 고려하여 Table 1과 같이 discriminator 네트워크를 설계하였다. 이때, Conv, LReLU, BN, 그리고 FC는 각각 콘볼루션, leaky ReLU, batch normalization, fully connected layer를 의미한다. Table 1의 discriminator는 블러 영상 개선 네트워크의 결과 영상 또는 정답 영상을 입력으로 하며, Wasserstein GAN [37]을 참조하여 다음 비용 함수로 학습된다.

Table 1. The architecture of discriminator network for (7)

Layer block	Weight dimension (stride)
Conv + LReLU	64×3×3×3 (1)
Conv + LReLU + BN	64×64×4×4 (2)
Conv + LReLU + BN	128×64×4×4 (2)
Conv + LReLU + BN	256×128×4×4 (2)
Conv + LReLU + BN	512×256×4×4 (2)
Conv + LReLU + BN	512×512×4×4 (2)
FC	1×512×4×4 (1)

$$L_D = D(S_{\lfloor M/2 \rfloor}) - D(G(B)), \tag{7}$$

이때, D와 G는 각각 discriminator 및 블러 영상 개선 네트워크를 나타낸다. (5)에서 L_{adv}가 개선 영상과 정답 영상의 구분 안 되게 학습하기 위해서는 다음의 비용 함수를 최소화해야 한다.

$$L_{adv} = -D(G(B)), \tag{8}$$

단일 블러 영상으로부터 옵티컬 플로우 추정 네트워크를 학습하기 위해서는 먼저 플로우에 대한 정답을 알아야 한다. 그러나 일반적으로 이러한 정답은 존재하지 않는다. 본 논문에서는 $S[0]$와 $S[M-1]$ 두 영상을 이용하여 최신의 성능을 보이는 FlowNet 2.0 [38] 기술을 이용하여 플로우에 대한 Pseudo 정답 ($\hat{F}_{0\to M-1}$와 $\hat{F}_{0\leftarrow M-1}$)을 먼저 계산하였다. Fig. 3은 제안하는 다중 학습 구조를 학습하기 위한 입력 영상 및 블러 개선과 플로우 추정에 대한 정답 영상의 예시이다. 이를 이용하여 L_{flow}는 다음과 같이 설계되었다.

$$L_{flow} = \| F_{forward} - \hat{F}_{0\to M-1} \|^2 + \| F_{backward} - \hat{F}_{0\leftarrow M-1} \|^2 \tag{9}$$

이때 $F_{forward}$와 $F_{backward}$는 옵티컬 플로우 추정 네트워크로 추정된 forward 및 backward 플로우를 나타낸다.

5. 실 험

5.1 학습 데이터 및 실험 환경

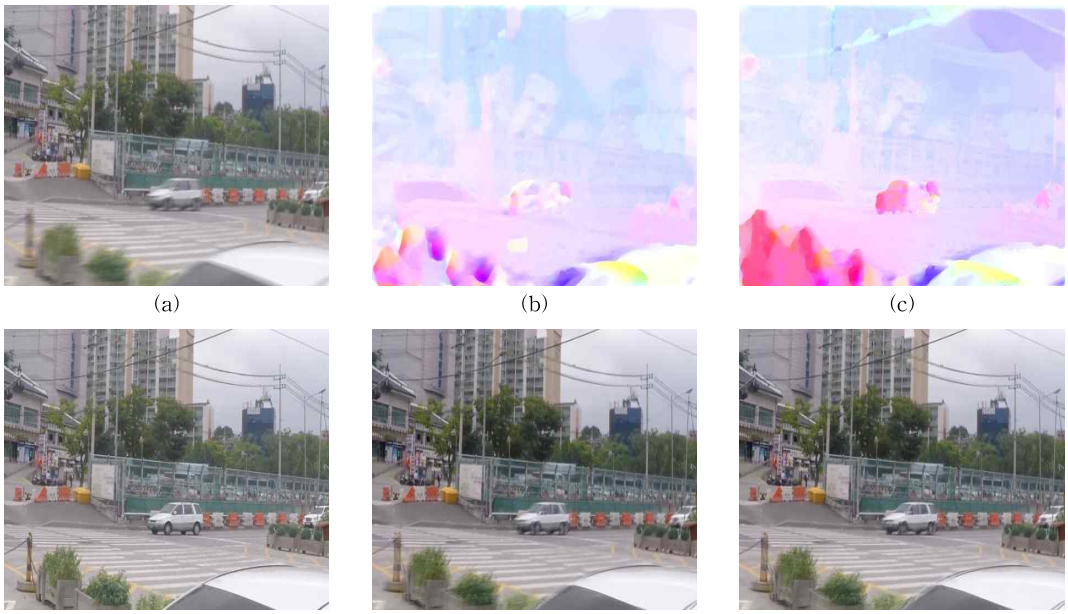
(a) (b) (c)

Fig. 3. (a) Blurry image, (b) Ground-truth for deblurring, and (c) Ground-truth for optical flow estimation.

본 논문에서는 객체/카메라 움직임에 의한 자연스러운 비균일 블러 영상을 이용하여 네트워크를 학습하기 위하여 GoPro 데이터베이스 [14]를 사용하였다. 이 데이터베이스는 2,103 쌍의 학습 데이터 및 1,111의 테스트 데이터로 구성되어 있으며, 각 영상은 1280×720의 해상도를 갖는다. 제안하는 네트워크를 학습하기 위하여 원본 영상으로부터 128×128 크기의 패치 80,000개를 랜덤으로 추출하였다. 모든 실험은 TensorFlow 라이브러리를 이용하여 학습되었으며, 12GB의 NVIDIA Titan GPU를 사용하였다. 전체 비용 함수 L_t는 ADAM solver [39]를 이용하여 최적화되었다 ($\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$). 배치 크기는 16으로 고정하여 총 100 Epoch를 학습하였다. 초기 러닝 비율은 10^{-4}이며, 20 epoch마다 러닝 비율을 반으로 감소시켰다. 모든 콘볼루션 레이어의 초기 파라미터는 Gaussian 분포를 갖게 하였다.

5.2 Ablation study

제안하는 다중 학습 구조의 효과를 확인하기 위하여 각각의 네트워크를 따로 학습한 것과 제안하는 다중 학습 구조를 학습한 결과를 비교하였다. Fig. 4는 이러한 Ablation study 실험 결과를 보여준다.



(a) (b) (c)

(d) (e) (f)

Fig. 4. Experimental results for ablation study: (a) Blurry image, (b) optical flow without deblurring netowork, (c) deblurry image without flow estimation network, (d) ground-truth shparp image, (e) optical flow via our multi-task, and (f) deblurry image via our multi-task.

Fig. 5. Visual comparison on dynamic blurry image. (a) Blurry image, (b) ground-truth, (c) Sun 등[17], (d) Nah 등[14], (e) DeblurGAN [15], and (f) Ours.

Table 2. Quantitative deblurring comparison on the GoPro dataset [14]

Measure	Sun 등[17]	Nah 등[14]	DeblurGAN [15]	Ours
PSNR	24.6	28.3	28.7	30.1
SSIM	0.842	0.916	0.958	0.969
Runtime	20 min	4.33 s	0.85 s	0.64 s

Fig. 4-(b)와 Fig. 4-(e)의 비교를 통해 단일 블러 영상으로부터 플로우 정보를 추정하는데 블러 개선 네트워크가 성능 향상에 도움을 주는 것을 확인할 수 있다. 마찬가지로 모션 정보 추정을 통해 블러 영상 개선의 성능이 향상된다는 것을 Fig. 4-(c)와 Fig. 4-(f)의 비교로 확인할 수 있다.

5.3 실험 결과 및 성능 비교

GoPro 데이터베이스 [14]는 1,111장의 테스트 영상을 포함하며, 이는 다양한 장소의 영상으로 구성되어 있다. 제안하는 다중 학습 구조는 위 테스트 데이터를 이용하여 최신의 블러 제거 기술 과 성능을 비교하였다. Fig. 5는 주관적 성능 평가로 제안하는 기술이 최신의 CNNs 기반 블러 제거 기술들보다 가시적으로 향상된 결과를 보이는 것을 확인할 수 있다. 원 영상과의 객관적 성능 비교를 위하여 품질 비교로 자주 사용되는 PSNR (Peak signal-to-noise ratio)와 영상의 구조적 유사성을 나타내는 SSIM (Structural similarity)를 사용하였다(Table 2). 객관적 성능 평가 또한 제안하는 기술이 가장 좋은 성능을 보이는 것을 확인할 수 있다. 마지막으로 각 알고리즘의 연산 속도를 비교하였다. 위 속도은 GPU 연산만을 포함한다. 제안하는 구조가 기존의 기술보다 더욱 빠른 속도로 블러 제거를 수행하는 것을 확인할 수 있다. 커널 추정 기반 방식인 Sun 등[17]의 기술은 개선 영상을 추정하기 위해 많은 반복 연산을 수행해야 한다. Nah 등[14]의 기술은 멀티스케일 기반의 구조로, 많은 콘볼루션 연산이 필요하다. 제안하는 다중학습 구조에 의해, DeblurGAN [15] 대비 적은 콘볼루션만 사용하여도 더 좋은 성능 및 연산 시간을 얻을 수 있다.

6. 결 론

본 논문에서는 다중 학습 구조를 통한 단일 블러 영상에서 블러 영상 개선 및 모션 정보 추정 기술을 제안하였다. 두 테스크는 서로 물리적으로 상호보완적인 특성을 갖고있으며, 네트워크 파라미터 공유를 통해 두 테스크의 성능을 향상시킬 수 있다. 다양한 Ablation 실험은 제안하는 다중 학습 구조의 효율성을 증명한다. 또한, 최신의 CNNs 블러 영상 개선 기술과의 비교를 통하여 제안하는 기술이 주관적/객관적 평가에서 가장 좋은 성능을 보임을 확인하였다.

REFERENCE

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778, 2016.

[2] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431-3440, 2015.

[3] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks," *Advances in Neural Information Processing Systems*, pp. 91-99, 2015.

[4] R. Fergus, B. Singh, A. Hertzmann, S.T. Roweis, and W.T. Freeman, "Removing Camera Shake from a Single Photograph," *Association for Computiong Machinery Transactions on Graphics*, Vol. 25, No. 3, pp. 787-794, 2006.

[5] D. Krishnan and R. Fergus, "Fast Image Deconvolution using Hyper-Laplacian Priors," *Advances in Neural Information Processing Systems*, pp. 1033-1041, 2009.

[6] D. Krishnan, T. Tay, and R. Fergus, "Blind Deconvolution Using a Normalized Sparsity Measure," *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 233-240, 2011.

[7] L. Xu, S. Zheng, and J. Jia, "Unnatural L0 Sparse Representation for Natural Image Deblurring," *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1107-1114, 2013.

[8] T.H. Kim, B. Ahn, and K.M. Lee, "Dynamic Scene Deblurring," *Proceeding of IEEE International Conference on Computer Vision,* pp. 3883-3891, 2013.

[9] T.H. Kim and K.M. Lee, "Segmentation-free Dynamic Scene Deblurring," *Proceeding of IEEE International Conference on Computer Vision,* pp. 2766-2773, 2014.

[10] H. Kim, K. Kwon, and J. Kim, "Adaptive Unsharp Masking Filter Design based on Multi-scale Retinex for Image Enhancement," *Journal of Korea Multimedia Society*, Vol. 21, No. 2, pp. 108-116, 2018.

[11] L. Xu, S.J. Ren, C. Liu, and J. Jia, "Deep Convolutional Neural Network for Image Deconvolution," *Advances in Neural Information Processing Systems*, pp. 1790-1798, 2014.

[12] O. Kwon, "Face Recognition based on Super-resolution Method Using Sparse Representation and Deep Learning," *Journal of Korea Multimedia Society*, Vol. 21, No. 2, pp. 173-180, 2018.

[13] J. Cho, H. Jang, N. Ha, S. Lee, S. Park, and K. Sohn, "Deep Unsupervised Learning for Rain Streak Removal Using Time-varing Rain Streak Scene," *Journal of Korea Multimedia Society*, Vol. 22, No. 1, pp. 1-9, 2019.

[14] S. Nah, T. Kim, and K.M. Lee, "Deep Multi-scale Convolutional Neural Network for Dynamic Scene Deblurring," *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3883-3891, 2017.

[15] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks," *Proceeding of IEEE Conference on Computer Vision and Pattern*

Recognition, pp. 8183-8192, 2018.

[16] C.J. Schuler, M. Hirsch, S. Harmeling, and B. Scholkopf, "Learning to Deblur," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 38, No. 7, pp. 1433-1451, 2016.

[17] J. Sun, W. Cao, Z. Xu, and J. Ponce, "Learning a Convolutional Neural Network for Non-uniform Motion Blur Removal," *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 769-777, 2015.

[18] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, et al., "Generative Adversarial Nets," *Advances in Neural Information Processing Systems*, pp. 2672-2680, 2014.

[19] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia, "Scale-recurrent Network for Deep Image Deblurring," *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8174-8182, 2018.

[20] J. Zhang, J. Pan, J. Ren, Y. Song, L. Bao, R.W. Lau, et al., "Dynamic Scene Deblurring Using Spatially Variant Recurrent Neural Networks," *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2521-2529, 2018.

[21] T. Brox and J. Malik, "Large Displacement Optical Flow: Descriptor Matching in Variational Motion Estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 3, pp. 500-513, 2011.

[22] C. Liu, J. Yuen, and A. Torralba, "SIFT Flow: Dense Correspondence Across Scenes and Its Applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 5, pp. 978-994, 2010.

[23] C. Barnex, E. Shechtman, D.B. Goldman, and A. Finkelstein, "The Generalized PatchMatch Correspondence Algorithm," *Proceeding European Conference on Computer Vision*, pp. 29-43, 2010.

[24] L. Xu, J. Jia, and Y. Matsushita, "Motion Detail Preserving Optical Flow Estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 34, No. 9, pp. 1744-1757, 2011.

[25] J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid, "Epicflow: Edge-preserving Interpolation of Correspondences for Optical Flow," *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1164-1172, 2015.

[26] J. Thewlis, S. Zheng, P.H. Torr, and A. Vedaldi, "Fully Trainable Deep Matching," *Proceeding of British Machine Vision Conference*, pp. 145.1-145.12, 2016.

[27] P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid, "DeepFlow: Large Displacement Optical Flow with Deep Matching," *Proceeding of IEEE International Conference on Computer Vision,* pp. 1385-1392, 2013.

[28] D. Gadot and L. Wolf, "Patchbatch: A Batch Augmented Loss for Optical Flow," *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4236-4245, 2016.

[29] A. Dosovitskiy, P. Fischer, E. Ilg, H. Philip, H. Caner, and G. Vladimir, "FlowNet: Learning Optical Flow with Convolutional Networks," *Proceeding of IEEE International Conference on Computer Vision,* pp. 2758-2766, 2015.

[30] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Deep End-to-end Voxel-to-voxel Prediction," *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 17-24, 2016.

[31] A. Ahmadi and I. Patras, "Unsupervised Convolutional Neural Networks for Motion Estimation," *Proceeding of IEEE International Conference on Image Processing,* pp. 1629-1633, 2016.

[32] D. Teny and M. Hebert, "Learning to Extract Motion from Videos in Convolutional Neural Networks," *arXiv Preprint arXiv:1601.07532*, 2016.

[33] A. Ranjan and M.J. Black, "Optical Flow Estimation Using a Spatial Pyramid Network," *arXiv Preprint arXiv:1611.00850*, 2016.

[34] H. Noh, S. Hong, and B. Han, "Learning De-

convolution Network for Semantic Segmentation," *Proceeding of IEEE International Conference on Image Processing,* pp. 1520-1528, 2015.

[35] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *arXiv Preprint arXiv: 1502.03167*, 2015.

[36] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-scale Image Recognition," *arXiv Preprint arXiv: 1409.1556*, 2014.

[37] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," *arXiv Preprint arXiv: 1701.07875*, 2017.

[38] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, and T. Brox, "Flownet2.0: Evolution of Optical Flow Estimation with Deep Networks," *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2462-2470, 2017.

[39] D.P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv Preprint arXiv: 1412.6980*, 2014.

정 형 주

2015년 연세대학교 전기전자공학과 학사
2015년~현재 연세대학교 전기전자공학과 석/박 통합과정
관심분야 : 영상처리, 영상개선, 탐지/추적

장 현 성

2007년 부산대학교 컴퓨터공학과 학사
2009년 부산대학교 컴퓨터공학과 석사
2011년~현재 LIG넥스원, 선임연구원
관심분야 : 딥러닝, 실시간 시스템, 영상처리

하 남 구

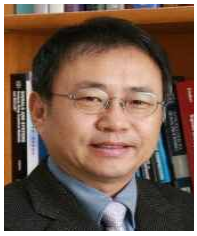
2008년 경북대학교 컴퓨터공학과 박사
2009년~현재 LIG넥스원, 수석연구원
관심분야 : 영상처리, 딥러닝, 감시정찰 시스템

연 윤 모

2006년 경북대학교 전자전기컴퓨터공학부 학사
2009년 경북대학교 전자전기컴퓨터공학부 석사
2009년~현재 LIG넥스원, 선임연구원
관심분야 : 영상처리, 임베디드 시스템, 실시간 시스템

권 구 용

2014년 한국 항공대학교 학사
2014년~현재 LIG넥스원, 선임연구원
관심분야 : 영상처리, 임베디드 시스템



손 광 훈

1983년 연세대학교 전자공학과 학사
1985년 미국 미네소타대학교 전자공학과 석사
1991년 미국 노스캐롤라이나 주립대학교 전기·컴퓨터공학과 박사
1988년~1992년 North Carolina A&T State University, Image Proc. & Computer Vision Lab., Research Associate
1992년~1993년 한국전자통신연구소 위성통신기술연구단, 선임연구원
1994년 Georgetown University MRI Center, Post Doctoral Fellow
1995년~현재 연세대학교 전기전자공학부 교수
2002년~2003년 Nanyang Technological University, School of Computer Engineering, Visiting Professor
2009년~2010년 국방과학연구소 전문위원
관심분야 : 3차원 영상 신호 처리, 컴퓨터 비전