

DenseFuse 기반의 광대역 영상 합성 학습을 위한 효과적인 데이터 분류 및 가중치 학습 모델

임찬기*¹, 손동민*², 권혁주*³, 이성학*⁴

Effective Data Classification and Weighted Learning Model for Wideband Image Synthesis Training Based on DenseFuse

Chan-Gi Im*¹, Dong-Min Son*², Hyuk-Ju Kwon*³, and Sung-Hak Lee*⁴

이 논문은 2021년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임
(NRF-2021R1I1A3049604)

요 약

본 논문은 기존의 다해상도 기반 가시광 및 근적외선 영상 합성 방법의 느린 처리 속도를 개선하기 위해 합성곱 신경망 기반 영상 합성 방법 중 하나인 DenseFuse를 활용하여 빠르게 합성하는 기법을 제시한다. 제안 방법은 효과적인 학습을 위해 래스터 스캔 방식의 알고리즘을 적용하여 가시광선 및 근적외선 데이터셋을 확보하고 휘도값과 분산값을 이용한 데이터셋 분류 방법을 제시한다. 또한, 합성 레이어에서 가중치 학습을 통한 효과적인 특징맵 합성 방법을 제시하였다. 제안 방법은 다해상도 기반 영상 합성 방법으로서 우수한 영상 합성 결과를 보였고 기존의 학습기반 영상 합성방법보다 가시성이 우수한 결과영상을 나타내었다. 제안 방법은 목표영상 생성에 사용된 다해상도 기반 영상 합성방법과 비교하여 처리시간을 1/3 이하로 단축시킴으로써 처리속도에서 강점을 보인다.

Abstract

In this paper, to improve the slow processing speed of the multi-resolution based visible and NIR(Near-infrared) image synthesis method, we present a fast synthesis method using DenseFuse, one of the CNN(Convolutional Neural Network)-based image synthesis methods. The proposed method applies a raster scan algorithm to secure visible and NIR datasets for effective learning, and presents a dataset classification method using luminance and variance. Also, in this paper, a method for synthesizing a feature map in a fusion layer is presented and compared with the method for synthesizing a feature map in other fusion layers. The proposed method learns the superior image quality of the multi-resolution based image synthesis method, and shows a clear synthesized image with better visibility than other existing learning-based image synthesis methods. Compared with the multi-resolution based image synthesis method used as the target image, the proposed method has the advantage in processing speed by reducing the processing time to 3 times or more.

Keywords

image fusion, deep learning, supervised learning, infrared image, visible image

* 경북대학교 IT대학 전자공학부(*⁴ 교신저자)
- ORCID¹: <https://orcid.org/0000-0002-0177-7634>
- ORCID²: <https://orcid.org/0000-0003-3007-8012>
- ORCID³: <https://orcid.org/0000-0002-7665-4049>
- ORCID⁴: <https://orcid.org/0000-0002-1030-381X>

• Received: Mar. 02, 2022, Revised: Jun. 27, 2022, Accepted: Jun. 30, 2022
• Corresponding Author: Sung-Hak Lee
School of Electronics Engineering, Kyungpook National University, 80
Daehak-ro, Buk-gu, Daegu, 41566, Korea
Tel.: +82-53-950-7216, Email: shak2@ee.knu.ac.kr

1. 서 론

CCTV, 로봇비전 및 자율주행차량의 영상은 신뢰성이 우선되고 객체 인식 기술이 온전히 수행되기 위해서 어떠한 상황, 공간 및 조명 상황에서도 오류가 없는 영상 정보를 제공해야 한다. 일반적으로 주간의 영상 센서는 가시광 영상 촬영을 위해 적외선 차단 필터(Hot mirror filter)를 사용하여 적외선 신호를 차단한다. 적외선 광 유입은 색 정보를 왜곡시키고, 낮 시간대의 강한 광량 때문에 적외선 광은 영상을 쉽게 포화시킬 수 있다[1][2]. 그러나 적외선 영상은 투과성과 회절성이 좋은 적외선광 특징으로 인해 가시광 영상에서 표현되지 않는 세부 정보를 포함하고 있으므로 정보 융합을 통해 영상의 가시성을 향상시킬 수 있다. 특히 안개, 연무 등이 낀 경우, 적외선은 가시광선 보다 입자에 대한 침투성이 강하기 때문에 적외선 영상으로부터 선명한 결과를 얻을 수 있다[3][4].

그림 1은 가시광 및 근적외선 영상을 이용한 영상 합성 결과를 보여준다. 근적외선 영상은 사물의 경계 성분을 많이 포함하고 있고 구름이나 안개에 가려진 사물은 근적외선으로 식별 가능하기 때문에 영상 개선에 있어서 근적외선 영상이 큰 역할을 한다[5]. 그러나 일반적으로 이러한 영상 합성은 정지 영상에서 이루어지며 각 영상의 국부 정보를 기반으로 합성되므로 처리 시간이 길어져 자율주행상황에서 객체 검출과 분류를 위한 추가적인 영상 합성 전처리는 적합하지 않다.



(a) 가시광 영상 (b) 근적외선 영상 (c) 합성 영상

그림 1. 가시광 및 근적외선 입력 및 결과 영상

Fig. 1. Visible light and near-infrared input and result images: (a) Visible image, (b) Near-infrared image, (c) Fused image

그러나 이러한 문제는 딥러닝 기반의 합성 기법을 통해 개선 될 수 있다[6]. 먼저 딥러닝 영상합성 학습을 위해서는 가시광 및 근적외선 영상 쌍의 데이터셋이 필요하다. 특히 시야와 전경이 동일한 영상 장면에 대한 개선도가 높은 영상 합성과 효과적인 딥러닝 학습 모델을 위해 많은 영상 쌍 데이터가 필요한데 충분한 촬영 영상을 구하기가 어렵다는 문제가 있다.

본 논문에서는 이러한 문제를 해결하기 위해 가시광 및 근적외선 영상 합성을 위한 효과적인 학습 데이터셋의 확보방법을 제시하고, 합성곱 신경망 기반의 영상 학습 방법 중 하나인 DenseFuse를 활용하여 합성 모델을 학습시키는 방법을 제안하였다. 최종적으로 검증 및 성능 평가를 위한 실험을 통해 기존의 학습 기반 영상 합성 방법과 비교하여 제안 방법이 우수함을 확인하였다.

II. 관련 연구

2.1 가시광 영상과 근적외선 합성

근적외선 파장 대역은 800~1500nm에 해당되고, 이는 시각적으로 인지할 수 있는 대역보다 높은 파장대역이다. 근적외선 영상에는 가시광 영상과 비교하여 물체의 경계 선명도가 높은 영역이 존재하고 가시광 및 근적외선 영상을 모두 촬영한 후 합성함으로써 영상의 가시성을 개선할 수 있다. 가시광 영상은 영상 합성 시 색성분 표현에 유리하고 근적외선 영상은 영상 전체의 경계와 질감이 표현되어 있어서 사물의 윤곽과 경계의 표현과 정확한 사물 판단에 유용하게 쓰일 수 있다. 따라서 안개나 연기가 많은 조건에서 근적외선 영상과 가시광 영상의 합성을 통해 선명하게 개선된 영상을 재생할 수 있다. 이러한 장점으로 인해 최근 다양한 가시광 영상과 근적외선 영상 합성 방법이 연구되고 있다[7]-[9].

Vanmali 등은 라플라시안-가우시안 피라미드 다해상도 분리와 국부 엔트로피 및 가시성 지수를 기반한 가중치 맵을 이용하여 가시광 영상과 근적외선 영상을 합성하였고, 색상 및 선명도 후보정의 과정을 거쳐 영상을 개선하였다[2]. 라플라시안-가우시안 피라미드는 입력 영상에서 다중 스케일 영상을

추출하는 다중 스케일 분해 알고리즘 중 하나로서 역방향 처리 시 두 개의 입력 피라미드 영상은 각각의 피라미드 레벨에 따라 별도로 합성된다. 이때 라플라시안 피라미드의 다중 스케일 영상은 입력 영상에서 차 영상과 가우시안 블러 업-다운 샘플링을 이용하여 얻을 수 있다.

그림 2는 가우시안 블러를 통한 라플라시안 피라미드 획득 방법을 보여준다. G 는 가우시안 필터를 의미한다[4].

2.2 딥러닝 기반 영상 합성

Prabhakar의 합성곱 신경망 기반의 딥러닝 모델은 삼(Siamese) 네트워크 아키텍처를 가지며, 인코더에 두 장의 합성곱 신경망 레이어, 디코더에 세 장의 합성곱 신경망 레이어를 구성하고 있다[5]. 두 가지 입력이 각각의 인코더로 들어가서 특징맵을 생성하고 이를 기반으로 합성하고 디코더에서 최종 영상을 출력한다. 이러한 합성 방법은 속도 면에서 우수한 성능을 보여주지만 네트워크 구조가 너무 간단하여 주요 특징들을 제대로 추출할 수 없다. 또한 인코딩 네트워크에 있는 마지막 레이어에 의해 계

산된 결과만 사용하여 중간 레이어에 있는 유용한 정보들이 소실된다는 문제가 있다.

DenseFuse는 가시광선 영상과 적외선 영상 합성곱에 적용된 신경망 모델로서 텐스블록(Dense block)이라는 캐스케이드(Cascade) 구조의 합성곱 신경망 레이어 집합을 사용하여 위의 정보 손실 문제를 해결한다.

그림 3은 텐스블록의 구조를 보여준다. Conv는 컨볼루션 블록(합성곱 신경망 레이어+활성함수)을 의미한다. 텐스블록은 3개의 컨볼루션 블록을 포함하고 각 컨볼루션 블록은 3x3 크기의 필터와 ReLu 활성화함수를 포함한다.

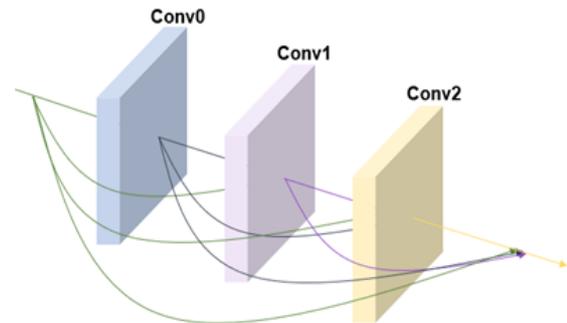


그림 3. 텐스블록의 구조
Fig. 3. Layout of dense blocks

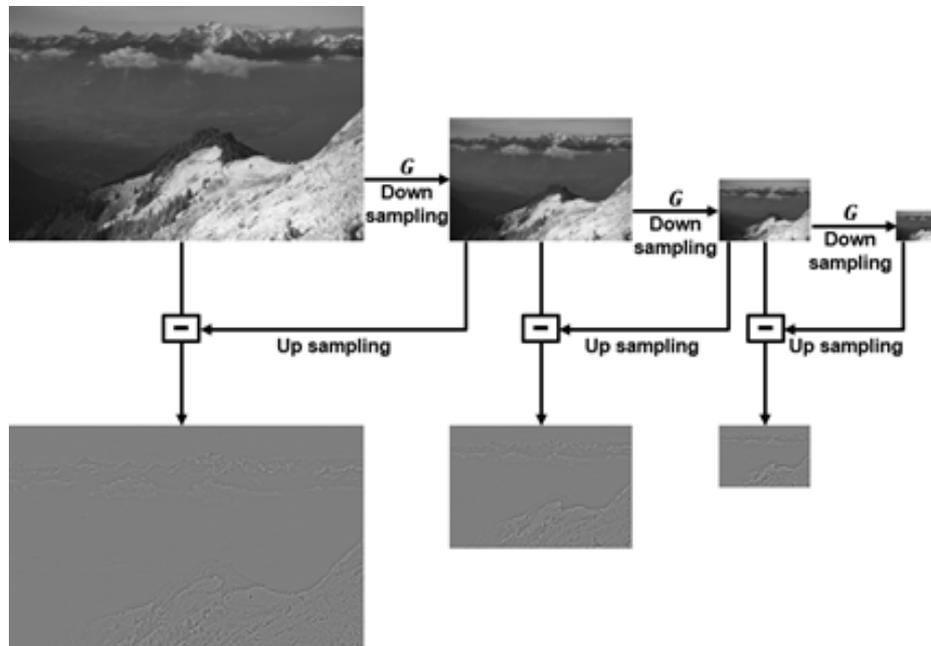


그림 2. 라플라시안 피라미드 처리 블록도
Fig. 2. Block diagram of Laplacian pyramid processing

그러나 DenseFuse 학습에서는 가시광선 영상과 적외선 영상 쌍의 데이터셋이 부족하여 가시광선 영상 데이터셋인 MS-COCO만을 학습에 사용하였다. 이로 인해 학습 시 인코더 네트워크에서 하나의 채널만을 사용함으로써 두 개의 채널을 사용하는 합성 아키텍처와 구조가 달라져 합성 시 영상의 품질을 보장하지 못한다는 문제가 있다[1][10].

III. 제안 방법

본 논문에서는 학습용 가시광 및 근적외선 영상 데이터셋 확보와 효과적인 영상 합성 학습을 위한 데이터 분류 및 가중치 학습 모델을 제안한다.

3.1 가시광 및 근적외선 영상 쌍 취득

먼저 목표 영상을 획득하기 위해 EPFL의 RGB-NIR 영상 데이터셋과 센서 다중 스펙트럼 영상 데이터셋(SSMID) 영상을 Vanmali의 합성 방법을 이용하여 합성한다[2][11][12]. 다음 목표 영상과 가시광 및 근적외선 영상 데이터셋으로부터 휘도채널 영상을 추출하여 학습용 입력영상 크기인 256×256 크기로 래스터 스캔(Raster scan) 알고리즘을 통해 일정한 간격으로 여러 장 잘라내어 확보한다.

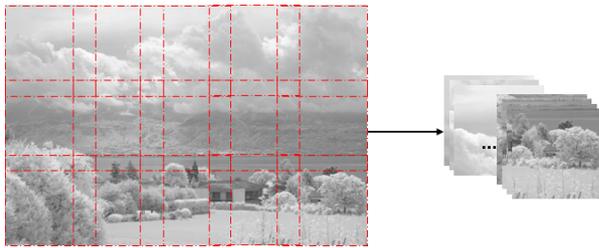


그림 4. 래스터 스캔 방식에 의한 영상 획득
Fig. 4. Image acquisition by a raster scan algorithm

그림 4는 래스터 스캔 방식의 알고리즘을 통해 1024×679 크기의 영상으로부터 입력 영상이 획득되는 과정을 나타낸다. 앞서 합성한 목표영상에 대해서도 제안 알고리즘을 적용시켜 학습할 목표영상을 획득한다.

3.2 효과적인 학습을 위한 영상 분류

취득된 영상은 비슷한 톤 영역 보다 차이가 큰 영역에서 합성 효과가 크다는 기존 합성 방법의 특징에 기반하여 휘도 차이 값과 분산 차이 값을 비교하여 분류된다.

가시광과 근적외선 영상 쌍의 차이 영상 및 평균 차이 값은 다음 식으로 구해진다. Dif_l 이 일정 값 이상인 가시광선 및 적외선 영상을 1차 선별한다.

$$I_{diff}(x,y) = |I_{vis}(x,y) - I_{nir}(x,y)| \quad (1)$$

$$Dif_l = \left[\sum_{x=0}^{N-1} \sum_{y=0}^{M-1} I_{diff}(x,y) \right] / NM \quad (2)$$

여기서 I_{vis} , I_{nir} 은 가시광 및 적외선 휘도 영상을 각각 나타내고 차이 영상을 I_{diff} 로 정의하였다. Dif_l 는 차이 영상의 휘도 평균값을 의미하고 NM 은 I_{diff} 의 전체 픽셀크기를 의미한다.

영상에서 국부 분산 값은 각 픽셀과 이웃하는 픽셀과의 휘도차이 분포를 나타낸다. 따라서 분산 영상은 영상에서 경계 성분 영역을 검출하기에 용이하다. 각 영상의 국부 영역 톤 차이로 개선 영상을 선별하기 힘들 경우 분산 영상의 차이를 통하여 개선 차이가 명확한 영상을 선별할 수 있다. 다음 식으로부터 블록기반 평균 연산자를 활용한 분산영상을 획득한다.

$$\mu = \left[\sum_{x'=-r}^r \sum_{y'=-r}^r I(x+x',y+y') \right] / (2r+1)^2 \quad (3)$$

$$V = \left[\left[\sum_{x'=-r}^r \sum_{y'=-r}^r I(x+x',y+y')^2 \right] / (2r+1)^2 \right] - \mu^2$$

여기서 μ 는 입력영상 I 의 국부 평균을 나타내고 V 는 각 픽셀에서 국부 분산 값을 나타낸다. r 은 블록의 크기를 의미하고 제안방법에서는 $r=2$ 를 사용하였다.

식 (3)를 통해 얻은 각각의 분산영상으로부터 식 (4) 및 (5)를 통해 분산 영상의 차 영상을 획득한다. 다음 Dif_l 과 마찬가지로 Dif_v 을 기준으로 분산영상의 차 영상의 평균값이 일정 값 이상인 가시광 및 적외선 영상을 분류한다.

$$V_{diff}(x,y) = |V_{vis}(x,y) - V_{nir}(x,y)| \quad (4) \quad L = 10^3 \times L_{ssim} + L_p \quad (6)$$

$$Diff_v = \left[\sum_{x=0}^{N-1} \sum_{y=0}^{M-1} V_{diff}(x,y) \right] / NM \quad (5) \quad L_p = \|O - T\|_2 \quad (7)$$

$$L_{ssim} = 1 - SSIM(O, T) \quad (8)$$

3.3 학습 모델

그림 5는 제안모델의 학습 구조를 나타낸다. 여기서 각 컨볼루션 레이어 하단의 숫자는 필터의 크기와 입력과 출력의 특징맵 수를 의미한다. 제안모델은 인코더, 합성 레이어, 디코더로 구성된다. 두 입력영상은 인코더에서 C11, DC11, DC21, DC31로 이루어진 채널 1과 C12, DC12, DC22, DC32로 이루어진 채널 2로 각각 들어간다. 각 채널에서 동일한 구조를 가지는 컨볼루션 레이어들이 각 입력에 대해 병렬적으로 연산하여 특징맵을 생성한다. 이때 DC11, DC21, DC31(DC12, DC22, DC32)는 캐스케이드 구조를 가지기 때문에 각 컨볼루션 레이어의 유용한 정보를 최대한 손실 없이 학습할 수 있다. 생성된 특징맵들은 합성 레이어에서 적절한 가중치가 곱해진 후 합성된다. 가중치는 학습을 통하여 정해진다. 합성된 특징맵은 디코더의 입력으로 들어가 디코더에서 총 네 장의 컨볼루션 레이어를 거쳐 최종 영상을 생성한다. 생성된 최종 영상을 준비된 목표 영상과 비교하여 손실 함수 L 이 최소가 되도록 인코더, 합성 레이어, 그리고 디코더를 학습한다.

다음 식에서 손실 함수 L 은 픽셀 손실 함수 L_p 와 구조적 유사도(SSIM) 손실 함수인 L_{ssim} 을 이용하여 계산된다.

여기서 O 와 T 는 각각 출력 영상과 목표 영상을 의미한다. 픽셀 손실 함수는 출력 영상과 목표 영상의 유클리드 거리(Euclidean distance)이다.

$SSIM(\cdot)$ 은 구조적 유사도 연산자를 나타내고, 두 영상 간 구조적 유사도를 비교한다[13].

3.4 영상 채널 합성

제안 모델의 영상 합성 방법을 그림 6에 나타내었다. 먼저, 합성할 영상을 CIELAB 색공간 변환을 통해 휘도 채널 l 과 색 채널 a, b 로 분리한다. CIELAB 색공간은 인간시각의 색인지 특성을 기반으로 있으며, 색 분리도가 우수하고 색성분 보존 및 보상을 위해 영상 톤맵핑 모델에 많이 활용된다[14].

가시광 영상의 색 채널 a_{vis}, b_{vis} 는 보존되어 합성 영상의 색 채널로 사용된다. 분리된 l_{vis}, l_{nir} 은 학습된 인코더의 각 채널의 입력으로 들어가 각 채널의 출력으로 나온 특징맵들은 합성 레이어에서 학습된 비율로 합성된다. 합성된 특징맵이 디코더의 입력으로 들어가 합성된 휘도영상인 l_{fused} 를 획득한다. 최종적으로 l_{fused} 에 보존되었던 a_{vis}, b_{vis} 를 병합하고 RGB 색공간 변환을 통해 컬러 영상을 획득한다.

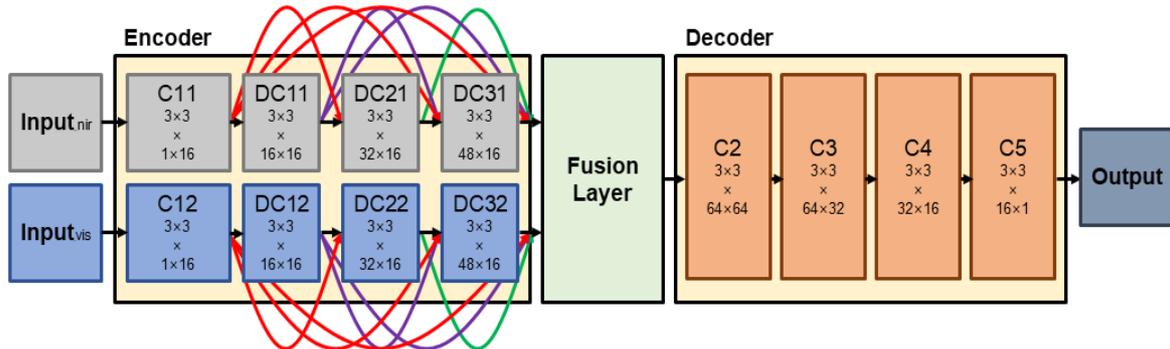


그림 5. 제안 모델의 학습 구조
 Fig. 5. Learning architecture of the proposed model

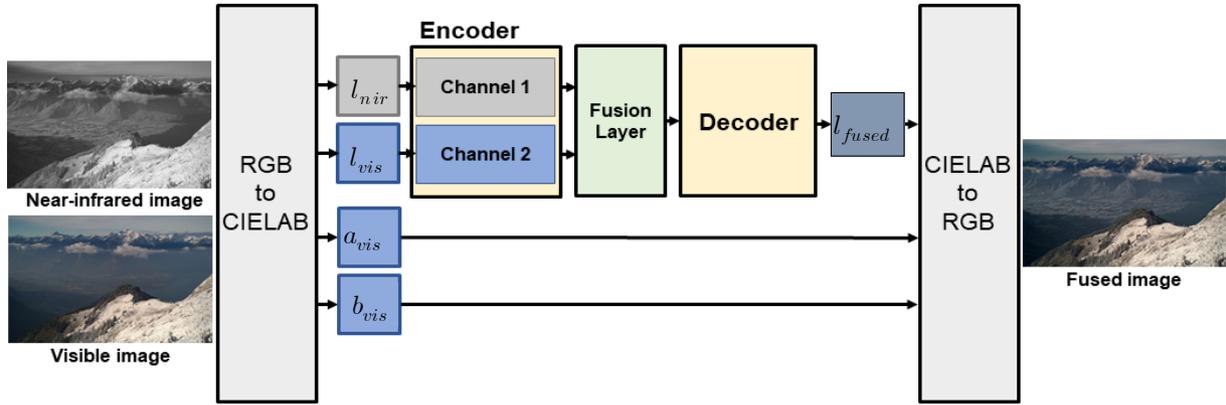


그림 6. 제안 모델의 영상 합성 방법
Fig. 6. Image fusion scheme of the proposed model

IV. 실험 결과

제안한 학습 모델의 성능을 확인하기 위해 여러 조건의 데이터셋을 제안모델을 통해 학습한 후 다양한 가시광선 및 근적외선 영상을 합성하여 SSIM 값을 통해 목표영상과 결과영상의 유사도를 비교하였다. 표 1은 각 데이터셋 별로 학습된 모델로부터 10장의 영상을 합성하고 합성영상과 목표영상 간 SSIM 값의 평균을 나타낸 것이다.

표 1. 데이터셋 별 SSIM 결과값 비교
Table 1. Comparison of SSIM values

Dataset	SSIM value
Lum_var_above	0.94288
Lum_above	0.94169
Lum_below	0.92600
Not_considered	0.93581

*lum_var_above*은 가시광과 근적외선 데이터셋 영상 간 휘도와 분산 차이의 평균값이 각각 13과 47 이상인 영상들로 선별한 데이터셋을 의미한다.

*lum_above*은 휘도 차이의 평균값만을 고려하여

선별한 데이터셋을 의미하고 *lum_below*은 휘도 차이의 평균값이 13 미만인 데이터셋이다. 마지막으로 *not_considered*은 휘도와 분산 차이의 평균값을 고려하지 않은 데이터셋이다. 결과를 통하여 *lum_var_above*으로 학습한 모델이 목표영상과 유사도가 가장 높다는 것을 확인할 수 있다.

또한, 합성 레이어에서 특징맵이 합성되는 비율에 따라 영상합성에 어떠한 영향을 주는지 확인하기 위해 합성 비율을 다르게 하여 결과 영상을 획득하고 기존의 가시광 및 적외선 영상 합성 방법인 Lowrank, DenseFuse, 그리고 Vanmali의 방법을 평가지표를 통해 비교하였다[15].

표 2와 그림 7은 합성방법 별로 총 10장의 가시광선 및 근적외선 합성 영상을 획득하고 평가지표를 통해 얻은 수치의 평균을 나타낸다. *Weighted_addition 1*은 제안 방법의 합성 레이어에서 근적외선 특징맵에 곱해지는 가중치만 학습을 한 모델을 나타내고, *Weighted_addition 2*은 적외선 특징맵과 가시광선 특징맵에 곱해지는 가중치를 각각 학습한 모델을 나타낸다.

표 2. 영상 정량 평가 점수
Table 2. Image quality metrics score

Method	LPC	FMI_pixel	S3	Qabf	
Lowrank[15]	0.92784	0.83760	0.13249	0.46567	
DenseFuse[1]	0.92169	0.85317	0.14407	0.53109	
Vanmali[2]	0.93204	0.85554	0.15531	0.56184	
Proposed model	Addition	0.93725	0.85979	0.14920	0.57635
	Weighted_addition1	0.93684	0.85849	0.14945	0.56836
	Weighted_addition2	0.93869	0.86039	0.15074	0.57138



그림 7. 영상 정량 평가지표 비교
Fig. 7. Comparison of image quality metrics

*Addition*은 특징맵에 곱해지는 가중치를 학습하지 않고 단순 합성 평균(Addition strategy)에 의해 합성된 모델이다. LPC와 S3은 영상의 선명도에 대하여 평가하고, FMI_pixel은 합성되는 두 입력영상의 정보를 얼마나 보존하였는지 나타낸다. Qabf는 영상의 품질에 대해 평가한다[16]-[19]. 여기서 *Weighted_addition* 2가 2가지 지표(LPC, FMI_pixel)에 대해 가장 좋은 값을 보였고, 나머지 지표(S3, Qabf)에서 두 번째로 좋은 값을 보였다. 이를 통해 제안 방법 중 *Weighted_addition* 2가 근소하게 영상 합성에 더 나은 결과를 나타낸다는 것을 확인할 수 있다. 전체적으로 제안 방법은 LPC, FMI_pixel 지표에서 기존방법 대비 1% ~ 3%의 미비한 개선도를 보였으나 S3, Qabf 지표에서는 8% ~ 24%의 정량지표 개선도를 보였다.

그림 8에서 그림 10까지 제안모델과 기존의 학습 기반 가시광선 및 적외선 영상 합성 방식의 결과 영상을 비교평가 하였다. 제안방법의 그림 8(e)와 목표 영상인 그림 8(d)는 그림 8(c)나 그림 8(b)에 비하여 그늘진 부분에서 영상의 디테일 정보가 잘 담긴 것을 확인할 수 있다. 특히 그림 8(b)는 그늘진 부분의 경계 성분이 표현 되지 않는다. 또한 그림 9(c)는 원거리의 산 부분에서 디테일이 잘 나타나지 않는 것을 확인할 수 있다.

특히 그림 9(b)에서는 산의 디테일 뿐만 아니라 영상 전반의 품질이 떨어지는 것을 확인할 수 있다. 반면 제안방법인 그림 9(e)는 또렷하고 품질이 우수한 영상을 보여준다.

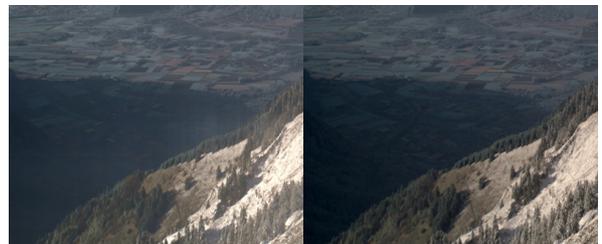
그림 10(b)와 그림 10(c)의 합성 결과는 산 경계와 나무 및 건물 등의 경계 표현이 떨어진다. 반면

제안 방법의 그림 10(e)는 목표 영상인 그림 10(d)과 같이 경계 표현이 잘되고 가시성이 우수하다.

표 3은 제안모델의 목표 영상 합성 방법인 Vanmali의 합성 방법과 제안 방법의 영상 합성 처리 시간을 나타낸다.



(a) 가시광 영상 및 NIR 휘도 영상
(a) Visible and NIR images



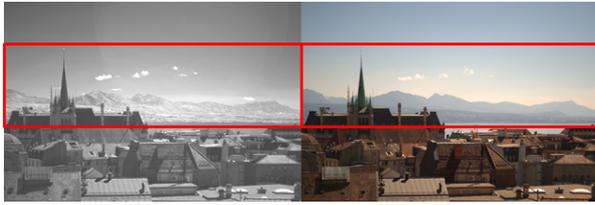
(b) Low rank 방법 결과 (b) Low rank
(c) Densefuse 방법 결과 (c) DenseFuse



(d) Vanmali 방법 결과(목표영상) (e) 제안 방법 결과
(d) Vanmali method(Target) (e) Proposed model

그림 8. 입력 및 결과 영상(1)

그림 8. Input and output images(1)



(a) 가시광 영상 및 NIR 휘도 영상
(a) Visible and NIR images



(b) Low rank 방법 결과
(b) Low rank



(c) DenseFuse 방법 결과
(c) DenseFuse



(d) Vanmali 방법 결과(목표영상)
(d) Vanmali method(Target)



(e) 제안 방법 결과
(e) Proposed model

그림 9. 입력 및 결과 영상(2)
Fig. 9. Input and output images(2)



(a) 가시광 영상 및 NIR 휘도 영상
(a) Visible and NIR images



(b) Low rank 방법 결과 (b) Low rank
(c) DenseFuse 방법 결과 (c) DenseFuse



(d) Vanmali 방법 결과(목표영상) (d) Vanmali method(Target)
(e) 제안 방법 결과 (e) proposed model

그림 10. 입력 및 결과 영상(3)
Fig. 10. Input and output images(3)

표 3. 제안모델과 목표영상방법 처리 시간 비교

Table 3. Comparison of processing time

Image resolution	Processing time(sec)	
	Vanmali model	Proposed model
580×320	0.3747	0.3589
1024×679	0.7111	0.3690
1920×1080	1.7711	0.5584

각각의 방법에 대해서 해상도 별로 10장의 영상을 합성하였고, 처리시간의 평균을 구하여 비교하였다. 해상도가 커질수록 딥러닝 기반의 합성방법이 규칙기반 합성방법보다 처리속도가 월등히 빠른 것을 확인할 수 있다.

V. 결론

본 논문에서는 딥러닝 기반의 가시광선 및 근적외선 영상 합성 방법을 이용하여 규칙 기반의 영상 합성 방법의 우수한 합성 품질을 보존하면서 동시에 처리속도를 단축시키는 방법을 제안하였다. 제안 방법은 효과적인 학습을 위한 데이터 셋 획득 방법과 분류 방법을 제시함으로써 규칙 기반 영상 합성 방법의 우수한 디테일 표현을 학습한다.

제안 방법은 정량적인 평가지표를 통해 기존의 학습기반의 합성방법과 비교하였으며 S3, Qabf 지표에서는 8% ~ 24%의 정량지표 개선도를 보였다.

그리고 결과 영상을 통한 시인성 비교에서도 제안 방법은 우수한 결과 영상을 보여줄 뿐만 아니라 목표 영상과 비교하여도 동등 이상의 합성 영상 품질을 보여준다. 또한 딥러닝 모델을 사용함으로써 계산량이 줄어들어 목표영상의 합성 방법보다 3배 이상 빠른 처리속도를 가진다. 이는 기존의 합성 방법들보다 동영상 합성에 더욱 적합한 방법으로 고려될 수 있음을 의미한다.

References

- [1] H. Li and X. J. Wu, "DenseFuse: A Fusion Approach to Infrared and Visible Images", *IEEE Transactions on Image Processing*, Vol. 28, No. 5, pp. 2614-2623, May 2019. <https://doi.org/10.1109/TIP.2018.2887342>.
- [2] A. V. Vanmali and V. M. Gadre, "Visible and NIR image fusion using weight-map-guided Laplacian-Gaussian pyramid for improving scene visibility", *Sādhanā : Academy proceedings in engineering sciences*, Vol. 42, No. 7, pp. 1063-1082, Jul. 2017. <https://doi.org/10.1007/s12046-017-0673-1>.
- [3] H. U. Lee, H. J. Kwon, and S. H. Lee, "Visible and NIR Image Fusion using the Subband Information of Wavelet Transform", *JKIIT*. Vol. 19, No. 3, pp. 77-86, Mar. 2021. <http://dx.doi.org/10.14801/jkiit.2021.19.3.77>.
- [4] D. M. Son, H. J. Kwon, and S. H. Lee, "Visible and Near-Infrared Image Synthesis Using PCA Fusion of Multiscale Layers", *Applied Sciences*, Vol. 10, No. 23, 8702, Dec. 2020. <http://dx.doi.org/10.3390/APP10238702>.
- [5] Q. Yan et al., "Cross-field joint image restoration via scale map", in *Proc. IEEE Int. Conf. Comput. Vis., The Chinese Univeristy of Hong Kong*, pp. 1537-1544, Oct. 2013.
- [6] K. R. Prabhakar, V. S. Srikar, and R. V. Babu, "DeepFuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs", 2017 *IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, pp. 724-732, Oct. 2017.
- [7] C. H. Son and X. P. Zhang, "Near-infrared fusion via color regularization for haze and color distortion removals", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 28, No. 11, pp. 3111-3126, Nov. 2018. <https://doi.org/10.1109/TCSVT.2017.2748150>.
- [8] C. H. Son and X. P. Zhang, "Near-infrared coloring via a contrast preserving mapping model", *IEEE Transactions on Image Processing*, Vol. 26, No. 11, pp. 5381-5394, Nov. 2017. <https://doi.org/10.1109/TIP.2017.2724241>.
- [9] L. Schaul, C. Fredembach, and S. Susstrunk, "Color image dehazing using the near-infrared", in *Proc. IEEE Int. Conf. Image Process., Cairo, Egypt*, pp. 1629-1632, Nov. 2009. <https://doi.org/10.1109/ICIP.2009.5413700>.
- [10] T. Y. Lin et al., "Microsoft COCO: Common objects in context", *European conference on computer vision*. Springer, Cham, pp. 740-755, Sep. 2014. https://doi.org/10.1007/978-3-319-10602-1_48.
- [11] X. Soria, A. D. Sappa, and A. Akbarinia, "Multispectral single-sensor RGB-NIR imaging: New challenges and opportunities", 2017 *Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*, Montreal, QC, Canada, pp. 1-6, Dec. 2017. <https://doi.org/10.1109/IPTA.2017.8310105>.
- [12] M. Brown and S. S. Susstrunk, "Multispectral SIFT for Scene Category Recognition", *International Conference on Computer Vision and Pattern Recognition (CVPR11)*, Colorado Springs, CO, USA, pp. 177-184, Jun. 2011. <https://doi.org/10.1109/CVPR.2011.5995637>.
- [13] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity", *IEEE Transactions on Image Processing*, Vol. 13, No. 4, pp. 600-612, Apr. 2004. <https://doi.org/10.1109/TIP.2003.819861>.
- [14] H. J. Kwon, S. H. Lee, G. Y. Lee, and K. I. Sohng, "Luminance adaptation transform based on

brightness functions for LDR image reproduction", Digital Signal Processing. Vol. 30, pp. 74-85, Jul. 2014. <https://doi.org/10.1016/j.dsp.2014.03.008>.

- [15] G. Liu and S. Yan, "Latent low-rank representation for subspace segmentation and feature extraction", 2011 International Conference on Computer Vision, Barcelona, Spain, pp. 1615-1622, Nov. 2011. <https://doi.org/10.1109/ICCV.2011.6126422>.
- [16] R. Hassen, Z. Wang, and M. M. A. Salama, "Image Sharpness Assessment Based on Local Phase Coherence", IEEE Transactions on Image Processing, Vol. 22, No. 7, pp. 2798-2810, Jul. 2013. <https://doi.org/10.1109/TIP.2013.2251643>.
- [17] M. Haghighat and M. A. Razian, "Fast-FMI: Non-reference image fusion metric", 2014 IEEE 8th International Conference on Application of Information and Communication Technologies (AICT), Astana, Kazakhstan, pp. 1-3, Oct. 2014. <https://doi.org/10.1109/ICAICT.2014.7036000>.
- [18] C. T. Vu, T. D. Phan, and D. M. Chandler, "A spectral and spatial measure of local perceived sharpness in natural images", IEEE Transactions on Image Processing, Vol. 21, No. 3, pp. 934-945, Mar. 2012. <https://doi.org/10.1109/TIP.2011.2169974>.
- [19] C. S. Xydeas and V. Petrović, "Objective image fusion performance measure", Electronics Letters, Vol. 36, No. 4, pp. 308-309, Mar. 2000.

저자소개

임 찬 기 (Chan-Gi Im)



2018년 3월 ~ 현재 : 경북대학교
IT대학 전자공학부 학부생
관심분야 : Deep Learning, Object
detection, 영상 처리, 영상 융합,
아날로그 회로, 디지털 회로

손 동 민 (Dong-Min Son)



2018년 9월 ~ 현재 : 경북대학교
대학원 전자전기공학부(박사과정)
관심분야 : 영상처리, HDR,
이미지센서, Deep learning,
Object detection

권 혁 주 (Hyuk-Ju Kwon)



2010년 2월 : 경북대학교 전자전기
컴퓨터학부(공학사)
2012년 2월 : 경북대학교
전자전기컴퓨터학부(공학석사)
2017년 9월 : 경북대학교
전자공학부(박사후 연구원)
2019년 8월 : 아진산업 선행연구팀

연구원

2019년 8월 ~ 현재 : 경북대학교 전자공학부 계약교수
관심분야 : Color Image Processing, Color Management
HDR 영상처리, 적외선 영상처리, 인공지능

이 성 학 (Sung-Hak Lee)



1997년 2월 : 경북대학교
전자공학과(공학사)
1999년 2월 : 경북대학교
전자공학과(공학석사)
1999년 2월 ~ 2004년 6월 : LG
전자 영상제품연구소 선임연구원
2008년 2월 : 경북대학교

전자공학과(공학박사)

2009년 8월 ~ 2017년 7월 : 경북대학교 IT대학 전자공학부
연구초빙교수

2018년 3월 ~ 현재 : 경북대학교 IT대학 전자공학부 교수
관심분야 : Color Image Processing, Color Appearance
Model, Color Management, HDR 영상처리, 영상융합,
인공지능영상처리