

# 전처리 기법에 따른 행동인식 모델 성능 차이 분석

박은수<sup>1)</sup>, 김승환<sup>1)</sup>, 굴람 무즈타바<sup>2)</sup>, 류은석<sup>1)</sup>  
성균관대학교<sup>1)</sup>, 가천대학교<sup>2)</sup>

## Analysis of Action Recognition Model Performance Differences According to Preprocessing Methods

Park, Eun-Soo<sup>1)</sup>, Kim, SeungHwan<sup>1)</sup>, Ghulam Mujtaba<sup>2)</sup>, Ryu, Eun-Seok<sup>1)</sup>  
Sungkyunkwan Univ.<sup>1)</sup>, Gachon Univ.<sup>2)</sup>

### 요 약

본 논문에서 행동 인식 데이터 셋의 변화 추세에 따라 현실적인 데이터 셋의 문제점인 화질 저하, 다양한 조도 등의 문제점을 고려하여 행동 인식 성능을 높이기 위해 전처리 기법을 적용한 실험을 수행한다. 적용하는 전처리 기법은 CLAHE, Bilateral filter, Median filter이며, 적용하는 옵션은 Augmentation이다. 결과로 CLAHE 기법을 적용했을 때 원본으로 실험한 정확도보다 약 4% 향상이 있으며 이에 대한 결과와 신경망 모델의 연관성에 관해 분석한다.

### 1. 서 론

최근 고성능 GPU의 사용으로 처리 가능한 연산량이 대폭 증가함에 따라, 많은 연산이 필요한 딥 러닝 관련 기술들이 연구가 활발히 진행되어 오고 있다. 이미지 처리에 특화된 Convolution Neural Networks(CNN)의 발달로 인하여 객체 인식, 행동 인식, 이미지 캡셔닝 등과 같은 이미지 프로세싱 연구가 진행되어 오고 있다. 이 중에서 행동 인식 관련 연구는 스켈레톤 데이터, 이미지 데이터, 시퀀셜 데이터 등 여러 방향을 통하여 활발히 진행되어 오고 있으며, 상당히 어려운 주제로 알려져 있다. 행동 인식은 영상 데이터를 입력으로 하기 때문에 많은 양의 데이터를 갖는다.

과거의 영상 관련 데이터 셋은 특정 상황에서 배우가 연기하여 같은 조도 및 앵글을 사용한 데이터 셋이 주를 이뤘었다. 그러나 현재 영상 관련 데이터 셋은 이미지의 크기가 다양하고 다양한 조도에서 촬영된 데이터 셋이 주를 이룬다. 따라서 조도 및 객체의 크기 등을 최대한으로 고려한 전처리 기법이 적용되면 모델의 성능이 증가한다. 본 논문에서 적용할 전처리 기법들은 다음과 같다. Histogram equalization, median filter, bilateral filter, augmentation. 본 논문에서 그림 1과 같이 위 나열한 기법들을 데이터 셋에 적용하여 어떤 전처리 기법이 행동 인식 테스트에 효과적인지 실험 및 분석한다.

본 논문의 구성은 다음과 같다. 2장에서 본 논문에서 적용할 전처리 기법을 설명한다. 3장에서 실험 후 결과를 통해 신경망 모델의 특성에 대해 분석한다. 4장에서 결론과 향후 연구를 제시하며 논문을 끝맺는다.

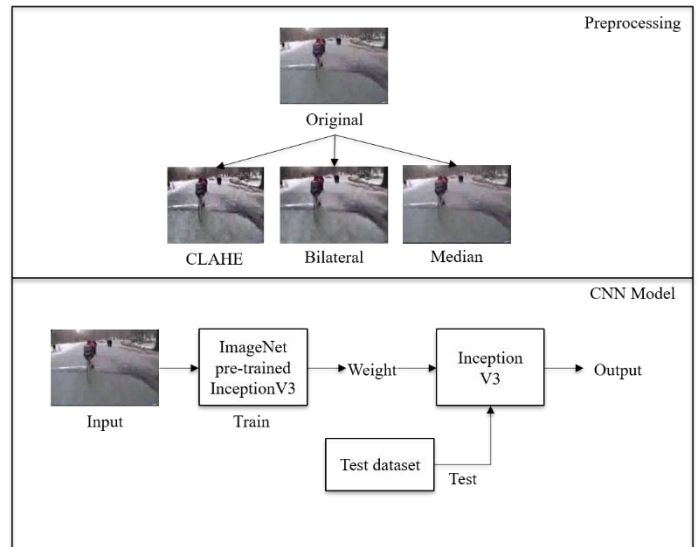


그림 1 제안하는 전처리 학습 기법

### 2. 관련연구

2 장에서 Histogram equalization, median filter, bilateral filter와 같은 전처리 기법에 대한 설명을 진행한다.

#### 2.1 Histogram Equalization

히스토그램 평활화는 영상의 그레이 스케일 (Grayscale) 값이 특정 값에 치우쳐져 있는 것을 넓고 고르게 퍼주는 이미지 처리법을 의미한다. 히스토그램 평활화의 공식은 (1)과 같다.

$$h(v) = \delta \left( \frac{cdf(v) - cdf_{min}}{(M \times N) - cdf_{min}} \right) \times (L - 1) \quad (1)$$

수식 (1)에서  $cdf$ 는 영상 내에서 해당 그레이 스케일 값을 가지는 픽셀의 빈도수를 표현한 확률밀도함수의 값들을 누적시킨 누적분포함수이다. 일반적으로  $cdf_{min}$ 을 분모, 분자에서 빼지 않고 진행하지만, 밝기 범위 전체 사용을 위해 빼고 진행한다. 분모  $M, N$ 은 영상의 가로, 세로 픽셀의 개수이다.  $cdf(v)$ 은 해당 그레이 스케일 값에서의  $cdf$ 값이다.  $L$ 은 최대 그레이 스케일 값이며  $cdf_{min}$ 은  $cdf$ 값 중 최솟값이다. 영상의 히스토그램이  $x$ 축이 명도,  $y$ 축이 빈도수라고 할 때 왼쪽으로 치우쳐져 있으면 해당 그레이 스케일 값이 전체적으로 작아서 영상의 밝기가 어두워진다. 반면 오른쪽으로 치우쳐져 있으면 전체적으로 커서 영상이 밝아진다. 히스토그램이 좁은 범위에 밀집되어 있으면 가장 어두운 명도와 가장 밝은 명도의 차이가 적기 때문에 명암 대비 (Contrast)가 좋지 않다. 히스토그램 평활화를 하게 되면 밀집된 그레이 스케일 값은 더 넓게 분산시키고, 흩어진 그레이 스케일 값은 밀집하게 되는 효과를 내게 된다. 이 과정을 통하여 명암 대비가 향상된다[1]. 히스토그램 평활화는 BBHE[1], RMSHE[2], DSHIHE[3], RSIHE[4]과 같이 영상 전체에 적용 또는 지역적 적용 이슈 및 재귀 방법에 따른 연구가 진행되어 왔다.

히스토그램 평활화를 영상 전체에 적용할 경우 원 영상의 특성에 따라서 명암 대비의 개선 효과가 다르게 나타난다. 그리고 지역적으로 밝기를 개선할 수 없기 때문에 명암 대비가 감소하거나 화질이 훼손된다[5,6]. 이를 개선하기 위하여 지역 히스토그램 평활화를 사용한다. 지역 히스토그램 평활화는 영상을 서브 블록으로 나누고 블록마다 히스토그램 평활화를 적용한다. 서로 다른 블록에서 각각 히스토그램 평활화가 적용되기 때문에 블록현상이 발생하는 문제점이 있다. 블록현상을 줄이기 위하여 명암 대비를 제한하는 Contrast limited adaptive histogram equalization (CLAHE)이 제안되었다[7].

## 2.2 Median Filter

Median filter는 중간 값 필터로 salt & pepper noise인 임펄스 잡음을 제거하는데 효과적이다. 통상적으로 이미지 프로세싱에서 윤곽선 감지 등의 처리를 수행 전 고성능 잡음 제거에 사용된다. 메디안 필터의 방식은 다음과 같다. (1) 선택된 픽셀을 기준으로 휴리스틱하게 생성한 마스크 생성 (2) 마스크 영역의 값을 크기 순으로 정렬 (3) 정렬된 값에서 중간값을 선택된 픽셀에 치환. 문제점으로 첫째, 연속된 픽셀에 발생한 노이즈를 제거하는데 문제가 있다. 둘째, 노이즈가 아닌 값(e.g. 윤곽선)도 값이 바뀌게 된다. 첫번째 문제는 마스크의 크기를 넓히면 해결된다. 그러나 두번째 문제점이 마스크의 크기를 넓힐 경우 문제점이 더 커지기 때문에 마스크의 크기의 절충점을 찾는 것이 중요하다.

## 2.3 Bilateral Filter

Bilateral filter는 경계를 보존하면서 노이즈를 감소시키는 블러 기법이다. 윤곽선이 무너지는 가우시안 필터의 문제점을 보완한다. 현재 픽셀과 이웃하는 픽셀 사이의 거리, 값의 차이를 동시에 가중치에 반영하여 윤곽선을 유지한다. 본 논문에서 CLAHE의 단점인 블록현상을 해결하기 위해 적용한다.

## 3. 전처리 적용 후 신경망 학습

3장에서 실험에 사용한 데이터 셋 및 실험 결과에 대한 분석을 진행한다. 실험은 앞서 소개한 전처리를 통해 어떤 전처리 기법이 성능을 향상시키는지 실험한다. 또한 성능이 향상된 이유에 관한 분석을 수행한다.

### 3.1 데이터 셋

행동 인식 모델 학습에 사용된 데이터 셋은 행동 인식에 많이 사용되는 UCF-101을 사용하였다. UCF-101 데이터 셋은 University of Central Florida에서 제작한 데이터 셋이다[8]. YouTube에서 다운받은 101가지 인간의 행동에 관련된 데이터 셋이 포함되어 있다. 총 13,320개의 비디오가 포함되어 있으며, 다양한 행동, 가변적인 카메라 움직임, 다양한 오브젝트 등을 포함하고 있다. 또한 여러 데이터 셋들은 현실과는 다른 준비된 영상들을 포함한 데이터 셋들이지만, UCF-101은 YouTube에서 다운로드 받아 직접 분류를 하였기 때문에 현실적인 요소를 포함하고 있다.

### 3.2 전처리 기법 적용 실험 및 결과 분석

본 논문에서 적용한 전처리 기법은 CLAHE, Median filter, Bilateral filter이다. Augmentation은 보편적으로 사용되는 데이터 셋 증강 기법이기 때문에 옵션으로 설정했다. Augmentation 기법은 기존 행동인식에서 주로 사용되는 Horizontal flip, Rescale, Random crop을 적용했다. 사용한 모델은 CNN 모델이다. 최근 경량화 모델에 관한 연구로 인해 CNN을 적용한 행동 인식 연구가 있다[9]. 실험 결과는 표 1과 같다. 표 1에서 None은 augmentation을 적용하지 않은 것을 뜻한다. High score는 loss값과 상관없이 가장 높은 정확도를 가질 때를 뜻한다. 표 1을 분석해볼 때 CLAHE와 Bilateral을 중요하게 볼 수 있다. CLAHE는 영상의 명암대비를 높여 윤곽선을 부각시키는 효과가 있지만 지역 히스토그램 평활화 기법이기 때문에 블록현상이 있어 이를 Bilateral filter로 완화시켰다. Bilateral의 결과로 볼 때 블러 처리가 된 영상은 학습 시 정확도가 높고 실험 시 정확도가 낮은 것으로 보아 오버피팅이 쉽게 일어난다고 분석할 수 있다. 반면 CLAHE 기법은 학습 시 크게 정확도가 높지 않았으나, 실험 시 정확도가 높은 것으로 보아 CLAHE를 적용한 결과가 신경망 모델에 긍정적인 효과를 미친 것을 알 수 있다. Augmentation을 적용하였을 때도 마찬가지로 CLAHE 기법을 적용하였을 때 성능이 향상되었다.

표 1. 전처리 기법을 적용한 행동인식 모델 실험 결과

	Method	Train accuracy (%)	Train loss	Test accuracy (%)	Test loss	High score
None	Original	87.7%	0.508	59.4%	1.600	59.4%
	CLAHE	87.0%	0.497	<b>63.8%</b>	<b>1.481</b>	<b>63.8%</b>
	Bilateral	<b>88.8%</b>	<b>0.433</b>	57.2%	1.674	58.4%
	Median	87.2%	0.467	58.1%	1.657	59.4%
Augmentation	Original	80.1%	0.743	59.1%	1.485	61.6%
	CLAHE	80.6%	0.742	<b>64.7%</b>	<b>1.399</b>	<b>64.7%</b>
	Bilateral	<b>82.0%</b>	<b>0.683</b>	54.4%	1.611	59.4%
	Median	79.5%	0.745	54.7%	1.764	55.9%

CLAHE는 앞서 언급하였듯 윤곽선을 부각하는 특성을 갖고 있는데 신경망 모델의 초기 레이어에서 윤곽 특징을 추출하고 후기 레이어에서 텍스처 특징을 추출하는 특성을 고려했을 때 CLAHE가 약한 윤곽 특징 추출 기능을 한 것이라 분석할 수 있다[10].

#### 4. 결론 및 향후연구

본 논문은 일반적으로 이미지 프로세싱에 사용되는 전처리를 통해 신경망 모델의 특징을 분석했다. 적용한 전처리 기법은 CLAHE, Bilateral, Median으로 CLAHE를 제외한 나머지 전처리 기법들은 성능 저하 현상이 발생했다. 이는 Bilateral과 Median이 블러 현상을 일으키므로 윤곽선을 보존하지 못하는 현상이 발생하기 때문이다. 반면 CLAHE는 블러현상이 다소 발생하더라도 윤곽선이 부각되어 성능 향상에 도움을 주는 것으로 볼 수 있다. 추후 연구로 본 실험을 통해 블러현상을 완화하면서 윤곽선을 부각할 수 있는 전처리 기법을 제안하여 모델 성능을 향상시킬 수 있다.

#### Acknowledgement

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터육성지원사업의 연구결과로 수행되었음" (IITP-2020-2017-0-01630)

#### 참 고 문 헌

[1] Kim, Y. T. (1997). Contrast enhancement using brightness preserving bi-histogram equalization. *IEEE transactions on Consumer Electronics*, 43(1), 1-8.

[2] Chen, S. D., & Ramli, A. R. (2003). Contrast enhancement using recursive mean-separate histogram equalization for scalable brightness preservation. *IEEE Transactions on consumer Electronics*, 49(4), 1301-1309.

[3] Wang, Y., Chen, Q., & Zhang, B. (1999). Image enhancement based on equal area dualistic sub-image

histogram equalization method. *IEEE Transactions on Consumer Electronics*, 45(1), 68-75.

[4] Sim, K. S., Tso, C. P., & Tan, Y. Y. (2007). Recursive sub-image histogram equalization applied to gray scale images. *Pattern Recognition Letters*, 28(10), 1209-1221.

[5] Gonzalez, R. C., Woods, R. E., & Eddins, S. L. (2004). *Digital image processing using MATLAB*. Pearson Education India.

[6] Kim, J. Y., Kim, L. S., & Hwang, S. H. (2001). An advanced contrast enhancement using partially overlapped sub-block histogram equalization. *IEEE transactions on circuits and systems for video technology*, 11(4), 475-484.

[7] Zuiderveld, K. (1994). Contrast limited adaptive histogram equalization. *Graphics gems*, 474-485.

[8] Soomro, K., Zamir, A. R., & Shah, M. (2012). UCF101: A dataset of 101 human actions classes from videos in the wild. *arXiv preprint arXiv:1212.0402*.

[9] Kozlov, A., Andronov, V., & Gritsenko, Y. (2020, March). Lightweight network architecture for real-time action recognition. In *Proceedings of the 35th Annual ACM Symposium on Applied Computing* (pp. 2074-2080).

[10] Dong, C., Loy, C. C., He, K., & Tang, X. (2015). Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2), 295-307.