

Video Super Resolution을 이용한 CCTV 영상 화질 개선

컴퓨터공학과
2015104221 정준현

개 요

1. 서론

1-1. 연구 배경

1-2. 연구 목표

2. 기존 연구

2-1. 기존 연구 1

2-2. 기존 연구 2

2-3. 기존 연구의 문제점

3. 프로젝트

3-1. 기존 연구와 차이점 및 해결방안

3-2. 프로젝트 내용

4. 진행 일정

5. 결론

참고문헌

1. 서론

저해상도의 Video를 상대적으로 고해상도인 video로 변환시키는 기술이 Video Super resolution이다. 고해상도의 video를 통해서 군사, 의학, 전기 전자 등 여러 가지 분야에서 중요하게 쓰일 수 있는 추가적인 정보를 얻을 수 있다. Video super resolution는 다양한 기술을 이용할 수 있다. 과거에는 일반적으로 Interpolation을 이용하여 super resolution을 구현했었다. 그러나 현재, 딥러닝의 부흥과 interpolation 방식의 저성능때문에 딥러닝의 여러 가지 인공지능망 모델을 통해 super resolution을 구현하려는 시도들이 늘어나고 있다. 본 프로젝트에서는 ‘SubPixel Convolutional neural network’를 이용하여 저해상도의 CCTV video를 고해상도로 바꾸는 것을 진행하고자 한다.

1-1. 연구배경

CCTV는 일생생활 전반에 걸쳐 교통 단속, 방법 및 범죄 해결, 재난 및 시설 감시등에 사용되고 있으나, CCTV를 통해 현장을 녹화한 영상을 살펴보면 해상도가 너무 낮거나 사람 또는 차량이 Blur하게 되어 원하는 내용을 파악하기 어려운 경우가 많다.

특히나 국보, 보물로 지정된 목조 문화재에 설치된 CCTV중 <그림 1>처럼 상당수의 해상도가 떨어져 재난 방지에 취약한것으로 나타나는 경우가 많다. 이는 CCTV 하드웨어를 교체하여 개선할 수 있으나, 이같은 저화질 CCTV를 교체하는 데는 대당 500만원 가량 정도가 든다고 한다. 이처럼 하드웨어적 해결방법이 없는 상태에서, 소프트웨어적 해결이 필요한 상황이다.



<그림 1. 41만 화소 이하 CCTV로 야간의 문화재를 촬영한 사진>

2. 연구목표

본질적인 목표는 이미 촬영된 저해상도 CCTV 영상에 대해서 고해상도로 변환하여 저해상도 영상에서는 알 수 없었던 유용한 정보를 고해상도의 영상에서 얻을 수 있도록 하는 것이 목표이다. 이에 따라 교통 단속, 방법 및 범죄 해결 등 여러 가지 분야에서 저

해상도의 CCTV때문에 발목을 잡혔던 부분을 해결하고자 한다.

이러한 본질적인 목표를 달성하기 위해서, 먼저 고해상도 뿐만 아니라 **blur**와 **noise** 제거를 통해 저해상도에서는 알 수 없었던 사물이나 객체가 고해상도에서는 그 정체를 알 수 있도록 하는 것이 중요하다. 대부분의 범죄, 교통 단속등 여러 가지 CCTV가 저해상도이기 때문에 발생하는 문제는 범인이나 사건의 중요한 부분이 **blur**하고 **noise**가 끼이기 때문에 그 정체를 식별할 수 없기 때문에 생긴다. 따라서 프로젝트를 진행하면서, 고해상도로 변환된 영상에 대해서 누구든지 그 영상 내의 모든 객체, 사물에 대해서 그것이 무엇인지 식별할 수 있도록 할 것이다.

2. 기존 연구

2-1. 기존 연구 1

‘Super Resolution Image Reconstruction by using Bicubic Interpolation’이라는 논문은 bicubic interpolation을 이용한 image super resolution을 구현한 논문이다. Interpolation이란 이미 알고 있는 값에 대해서 현재 알지 못하는 값을 예측하는 것이다. 여러 가지의 Interpolation 기법중에서 bicubic interpolation은 16 pixel (4x4)의 값을 고려하여 interpolation한다. Bilinear interpolation와 같은 다른 interpolation은 이보다 적은 픽셀을 고려하기 때문에, 속도 측면이 중요하게 고려되어지지 않아도 된다면 bicubic interpolation을 이용하는 것이 더 보편적이다.

일반적인 Bicubic interpolation의 방정식은 <방정식 1>와 같다.

$$p(x, y) = \sum_{i=0}^3 \sum_{j=0}^3 a_{ij} x^i y^j.$$

<방정식 1>

f, f_x, f_y, f_{xy} 와 같은 변수들은 (0, 0), (1, 0), (0, 1), (1, 1)에서 이미 알고 있다고 가정하자. 먼저 f 좌표 (0, 0)에 대해 4개의 좌표로 변환하기 위해, interpolated surface를 구해보면 <방정식 1>과 같다.

$$\begin{aligned}
1. & f(0, 0) = p(0, 0) = a_{00}, \\
2. & f(1, 0) = p(1, 0) = a_{00} + a_{10} + a_{20} + a_{30}, \\
3. & f(0, 1) = p(0, 1) = a_{00} + a_{01} + a_{02} + a_{03}, \\
4. & f(1, 1) = p(1, 1) = \sum_{i=0}^3 \sum_{j=0}^3 a_{ij}.
\end{aligned}$$

<방정식 2>

$p(x, y)$ 를 function value와 매칭시키는 것은 <방정식 2>와 같은 4개의 방정식을 내놓게 된다. 이와 같은 방식으로 f_x, f_y, f_{xy} 좌표에 대해 각각 4개의 좌표를 계산하는 방정식을 통해 총 16가지의 방정식을 산출하게 된다. 이를 역행렬 계산등의 과정을 거쳐 궁극적으로 16개의 a_{ij} 의 값을 구하게되면, 16개의 좌표를 구할 수 있게 된다. 이것이 bicubic interpolation의 핵심 logic이다. Bicubic interpolation은 다른 interpolation에 비교했을 때 <표 1> 과 같은 성능을 보였다고 한다.

Interpolation techniques	PSNR (in dB)	Time in Seconds
Nearest neighbor	26.05	2.4990
Bilinear	27.12	2.522
Bicubic Interpolation	27.32	4.599

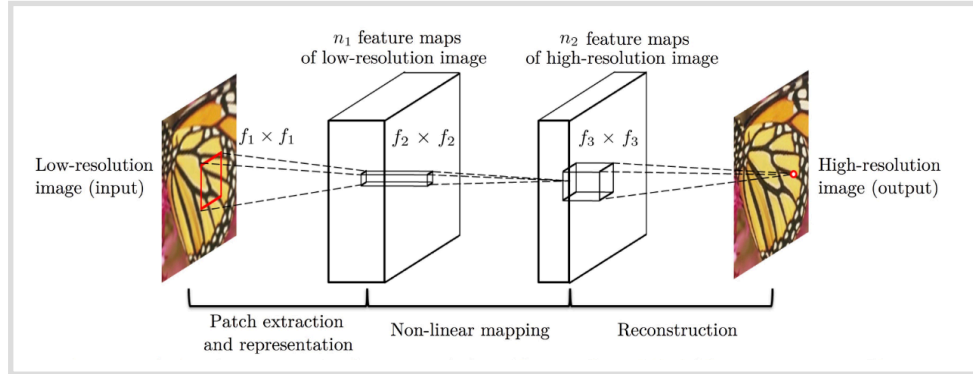
<표 1>

2-2. 기존 연구 2

‘Convolutional Neural Networks for Super-Resolution’이라는 논문에서 입력 저해상도 이미지는 Y, 복원한 출력 고해상도 이미지는 F(Y), ground truth 고해상도 이미지는 X로 표현하고 있다. 고해상도 이미지를 복원하는 mapping F는 아래의 세 가지 연산으로 구성된다.

- Patch extraction and representation: 저해상도 이미지 Y로 부터 patch 추출
 $\rightarrow F1(Y) = \max(0, W1 * Y + B1)$
- Non-linear mapping: 다차원 patch 벡터를 다른 다차원 patch 벡터로 mapping
 $\rightarrow F2(Y) = \max(0, W2 * F1(Y) + B2)$
- Reconstruction: 다차원 patch 벡터에서 최종 고해상도 이미지 생성
 $\rightarrow F(Y) = W3 * F2(Y) + B3$

W는 convolution 필터, B는 bias, *는 convolution operation을 의미한다. 이와 같은 세가지 연산을 Layer단위로 처리하는 CNN 구조(그림 2)를 이용하여 Super resolution하는 방법을 위 논문에서는 제안하였다.



<그림 2. Super resolution CNN 구조>

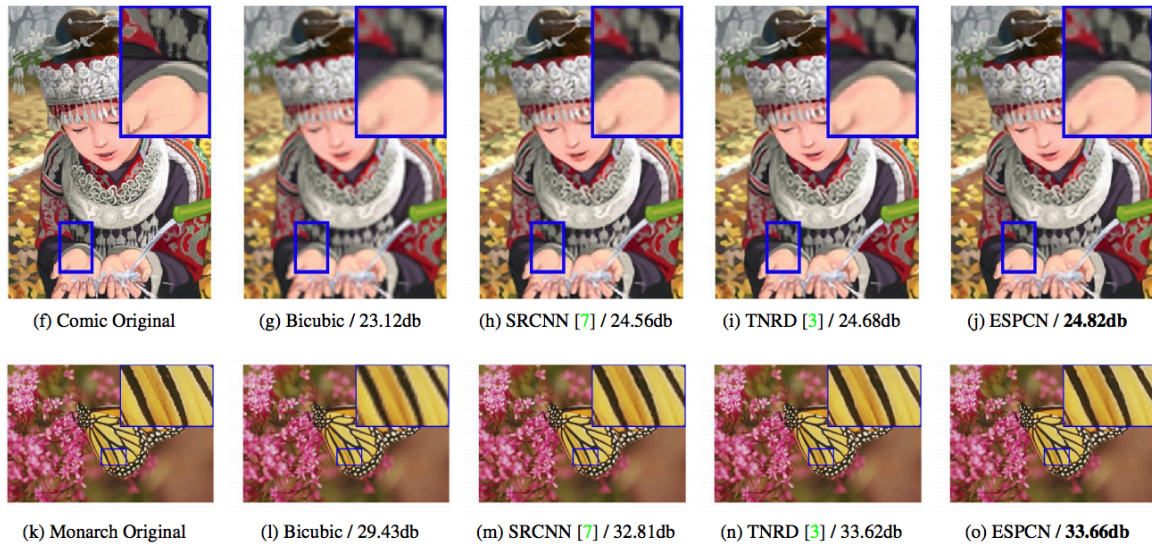
Loss 함수는 <방정식 3> 과 같이 Mean Squared Error (MSE)를 사용한다.

$$L(\Theta) = \frac{1}{n} \sum_{i=1}^n \|F(\mathbf{Y}_i; \Theta) - \mathbf{X}_i\|^2$$

<방정식 3>

2-3. 기존 연구의 문제점

Bicubic Interpolation, SRCNN(Super Resolution CNN)의 경우 최근에 나온 여러 가지 Super Resolution 기술보다 성능이 떨어지는 것이 가장 큰 문제점이다. <그림 3>에서 볼 수 있듯, Bicubic super resolution의 경우 PSNR의 값이 SRCNN을 이용한 경우보다 현저히 낮은 것을 볼 수 있으며, SRCNN또한 최근에 나온 ESPCN(Efficient Sub-Pixel Convolutional Neural Network)보다 PSNR의 값이 미세하게 작은 것을 볼 수 있다. 또한 실제 복원된 결과를 봤을 때, bicubic interpolation과 SRCNN의 경우보다 ESPCN의 경우가 더 잘 복원된 것을 볼 수 있다.



<그림 3. 실제 사진 및 그림에 대해서 다양한 super resolution 기술을 이용한 예 >

3. 프로젝트

3-1. 기존 연구와 차이점 및 해결방안

기존 연구는 image 혹은 video를 High Resolution으로 restoration하는 과정이 computation overhead가 매우 크다. 따라서 이에 대한 accuracy가 떨어져서 결과의 품질이 낮아질 수 밖에 없다. 또한, SRCNN의 경우 단순한 CNN의 layer가 3개정도로 쌓여서 Low resolution 이미지 혹은 비디오를 High resolution으로 transformation하게 된다. 이에 비해 ESPCN의 경우 Sub-pixel Convolution layer를 뒤서 pixel shuffling 을 하게 되는데, 이것이 SRCNN과 ESPCN의 큰 차이점이다.

본 프로젝트에서는 ESPCN 모델을 이용하여 ImageNet의 데이터셋을 가지고 학습하여, 저해상도의 CCTV 영상을 고해상도의 영상으로 변환하는 application을 설계 및 구현한다.

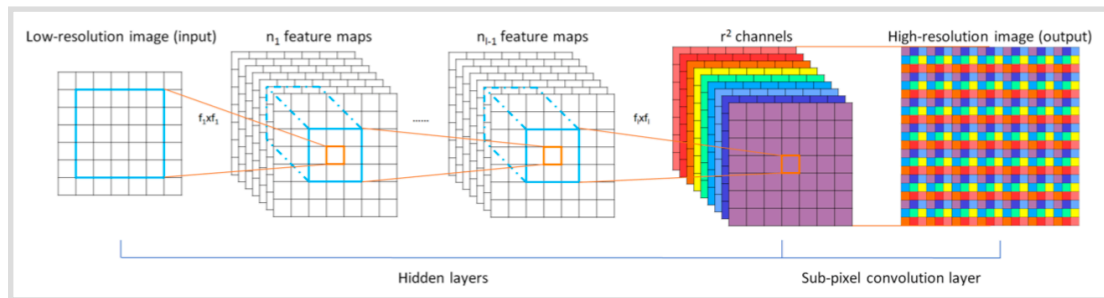
3-2. 프로젝트 내용

본 프로젝트에서 사용할 기술인 ESPCN의 구조는 <그림 4>와 같다. Low - Resolution Image에 대해서 Hidden CNN Layer를 거쳐 고도화된 feature maps를 추출하게 되고, 마지막에 sub-pixel convolution layer에서 upscaling하게 된다.

Low Resolution Image Input에 대해서 따로 전처리 하지 않고 그대로 ESPCN에 Input으로 넣어주면 되기 때문에, convolution filter의 size또한 줄어들게 되고 이는

computation overhead를 상당히 많이 줄여줄 수 있게 된다.

여러개의 Layer 각각마다 여러 개의 upscaling filter를 학습하게 되기 때문에, Low Resolution에서 High Resolution으로 mapping시키는 과정이 더 복잡하며 좋다. 이에 따라, reconstruction accuracy 즉, 결과의 품질이 더 좋아지게 된다.



<그림 4. ESPCN의 구조>

본 프로젝트에서는 ESPCNN모델을 구현한 후, Image Net으로 부터 랜덤하게 추출된 50,000개의 이미지를 학습 시키게 될 것이다. Upscaling filter, convolutional filter등 학습된 Weights 들을 이용하여 저해상도 CCTV 영상을 High Resolution Video로 변환하고자 한다.

4. 진행 일정

진행 주차	담당	내용
4주차	정준현	주제 확정 및 관련 논문 및 자료 조사
5주차~6주차	정준현	모델 설계, 데이터 수집 및 논문 초안 작성
7주차~10주차	정준현	설계 검증 및 초기 모델 구현
11주차~14주차	정준현	데이터 학습 및 테스트 데이터에 대해 결과 검증
15주차~16주차	정준현	어플리케이션 프로토타입 준비, 발표 준비

5. 결론

저해상도의 CCTV 비디오를 ESPCN을 이용하여 고해상도의 CCTV 영상으로 바꿔주는 Application을 구현할 것이다. 이 Application은 CCTV 영상이 저해상도임에 따라 발생하는 여러 가지 범죄, 방법등의 문제를 해결하는 데 있어서 방향성을 잡아줄 것이다.

참고 문헌

- Wenzhe Shi, Jose Caballero, Ferenc Husza'r, Johannes Totz, Andrew P. Aitken, Rob Bishop, Daniel Rueckert, Zehan Wang(2016), 「Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network」, Magic Pony Technology, Imperial College London.
- Prasanna S. Sane, Anil B. Gavade(2014), 「Super Resolution Image Reconstruction By Using Bicubic Interpolation」, Department of Electronics and Communication, KLS Gogte Institute of Technology.
- Chao Dong, Chen Change Loy, Kaiming He, Xiaoou Tang (2015), 「Image Super-Resolution Using Deep Convolutional Networks」,