

영상 관찰 모델을 이용한 예제기반 초해상도 텍스트 영상 복원

박 규 로[†] · 김 인 중^{††}

요 약

예제기반 초해상도 영상 복원(EBSR)은 고해상도 영상과 저해상도 영상간의 패치간 대응관계를 학습함으로써 고해상도 영상을 복원하는 방법으로, 한 장의 저해상도 영상으로부터도 고해상도 영상을 복원할 수 있는 장점이 있다. 그러나, 폰트의 종류나 크기가 학습 영상과 다른 텍스트 영상을 적용할 경우 잡영을 많이 발생시킨다. 그 이유는 복원 과정 중 매칭 단계에서 입력 패치들이 사전 내의 고해상도 패치와 부적절하게 매칭될 수 있기 때문이다. 본 논문에서는 이러한 문제점을 극복하기 위한 새로운 패치 매칭 방법을 제안한다. 제안하는 방법은 영상 관찰 모델을 이용하여 입력 영상과 출력 영상간의 상관 관계를 보존함으로써 잘못 매칭된 패치로 인한 잡영을 효과적으로 억제한다. 이는 출력 영상의 화질을 개선할 뿐 아니라, 다양한 종류 및 크기의 폰트를 포함한 대용량 패치 사전을 적용할 수 있게 함으로써 폰트의 종류 및 크기의 변이에 대한 적응력을 크게 향상시킨다. 실험에서 제안하는 방법은 폰트와 크기가 다양한 영상에 대하여 기존의 방법보다 우수한 영상 복원 성능을 나타내었다. 뿐만 아니라, 인식 성능도 88.58%에서 93.54%로 개선되어 제안하는 방법이 인식 성능의 개선에도 효과적임을 확인하였다.

키워드 : 예제기반 초해상도 영상 복원, 패치 검색 기준, 영상 관찰 모델, 문서 영상 복원

Example-based Super Resolution Text Image Reconstruction Using Image Observation Model

Gyu-Ro Park[†] · In-Jung Kim^{††}

ABSTRACT

Example-based super resolution(EBSR) is a method to reconstruct high-resolution images by learning patch-wise correspondence between high-resolution and low-resolution images. It can reconstruct a high-resolution from just a single low-resolution image. However, when it is applied to a text image whose font type and size are different from those of training images, it often produces lots of noise. The primary reason is that, in the patch matching step of the reconstruction process, input patches can be inappropriately matched to the high-resolution patches in the patch dictionary. In this paper, we propose a new patch matching method to overcome this problem. Using an image observation model, it preserves the correlation between the input and the output images. Therefore, it effectively suppresses spurious noise caused by inappropriately matched patches. This does not only improve the quality of the output image but also allows the system to use a huge dictionary containing a variety of font types and sizes, which significantly improves the adaptability to variation in font type and size. In experiments, the proposed method outperformed conventional methods in reconstruction of multi-font and multi-size images. Moreover, it improved recognition performance from 88.58% to 93.54%, which confirms the practical effect of the proposed method on recognition performance.

Keywords : Example-Based Super Resolution, Patch Matching Criterion, Image Observation Model, Text Image Reconstruction

1. 서 론

휴대용 카메라의 보급이 증가함에 따라 카메라 기반 문서 영상 인식 기술의 응용 분야도 넓어지고 있다. 그러나, 현재의 문서 인식 시스템들은 저화질의 카메라 영상에 대하여는 성능이 많이 저하된다[1-4]. 이러한 한계를 극복하기 위하여 초해상도 영상 복원을 저해상도 영상의 인식에 적용하기 위한 연구가 많이 수행되었고, 이들은 좋은 결과를 나타내었

* 이 논문은 2007년 정부(교육과학기술부)의 재원으로 한국연구재단(구 학술진흥재단)의 지원을 받아 수행된 연구임(KRF-2007-331-D00420).

※ 본 연구는 교육과학기술부와 한국산업기술재단의 지역혁신인력양성사업으로 수행된 연구결과임.

† 준 회원 : 한동대학교 정보통신공학과 석사과정

†† 종신회원 : 한동대학교 교수

논문접수: 2010년 6월 4일
수정일: 1차 2010년 7월 13일, 2차 2010년 7월 20일
심사완료: 2010년 7월 21일

<표 1> EBSR 시스템의 복원 결과(단일 폰트, 단일 크기로 학습)

		Tahoma	Times New Roman
19pt	입력영상	as	as
	EBSR 결과	as	as
13pt	입력영상	as	as
	EBSR 결과	as	as

다. 이러한 방법들 중 하나는 예제기반 초해상도 영상 복원 (example-based super-resolution: EBSR)이다[4, 5].

EBSR은 학습 단계에서 고해상도 학습 영상으로부터 패치들을 추출하여 패치 사전에 저장한다. 복원 단계에서는 패치 사전에 저장된 고해상도 패치들을 이용하여 저해상도 입력 영상의 세부 형태를 복원함으로써 고해상도 영상을 생성한다. EBSR은 한 장의 저해상도 영상으로부터도 고해상도 영상을 생성할 수 있을 뿐 아니라, 복원 성능도 우수하다[4, 5]. 그러나, 폰트 종류 및 크기가 학습 영상과 다른 텍스트 영상에 적용할 경우 잡영을 많이 발생시킨다.

<표 1>은 Freeman의 방법에 기반한 EBSR 시스템의 결과이다[4]. 패치 사전은 19pt Tahoma 폰트에 의해 학습되었다. 입력 영상의 폰트 종류 및 크기가 학습 영상과 동일한 경우에는 좋은 결과를 출력했으나, 그렇지 않은 영상에 대해서는 많은 잡영이 발생하였다. 이와 같은 문제는 [5]를 비롯한 다른 문헌에서도 보고된 바 있다.

이러한 문제를 극복하기 위해서는 패치 사전이 다양한 종류와 크기의 폰트들을 포함해야 한다. 그러나, 패치 사전을 확대하는 것만으로는 충분하지 않다. <표 2>는 11가지 종류와 18가지 크기의 폰트로 학습된 대용량 패치 사전을 이용한 EBSR 결과이다.

충분한 패치가 사전에 존재했음에도 불구하고, 오히려 <표 1>보다 더 많은 잡영이 발생하였다. 그 이유는 입력 패치들 중 일부가 부적절한 고해상도 패치와 매칭되었기 때문이다. EBSR 시스템은 저해상도 입력 패치를 패치 사전 내의 고해상도 패치와 매칭한 후 각 입력 패치와 매칭된 고해

<표 2> EBSR 시스템의 복원 결과 (다중 폰트, 다중 크기로 학습)

input images	asaSasas
EBSR results	asaSasas

상도 패치를 이용하여 영상의 세부 형태를 복원한다. 기존 연구에서는 패치 키를 이용한 매칭 방법이 많이 사용되었다 [4, 5]. 학습 단계에서 각 고해상도 패치에 검색을 위한 키를 지정한 후, 복원 단계에서 입력 패치의 키와 고해상도 패치의 키를 비교함으로써 매칭을 수행한다.

패치 키로는 화질 저하 시 잘 소실되지 않는 중간 주파수 (mid-frequency) 성분이 많이 사용된다[4, 5]. 그러나, 중간 주파수 성분은 고해상도 패치와의 상관 관계가 충분히 강하지 못하다. 중간 주파수 성분이 매우 유사하면서도 고주파 성분이 다른 패치가 다수 존재할 수 있기 때문이다. 고해상도 패치와의 상관 관계가 강하지 않은 키를 이용하여 패치를 매칭할 경우 각 입력 패치는 부적절한 고해상도 패치와 매칭될 수 있는데, 이러한 패치를 복원에 사용할 경우 잡영이 발생한다. 여러 연구에서 이러한 문제를 완화하기 위해서 인접한 고해상도 패치간의 관계를 매칭에 반영하였다[4, 5]. 그러나, <표 2>의 결과들은 이러한 방법이 충분하지 않다는 것을 보여준다.

본 논문에서는 텍스트 영상의 EBSR을 위한 새로운 패치 매칭 방법을 제안한다. 제안하는 방법은 영상 관찰 모델로부터 유도된 매칭 기준을 이용하여 패치를 매칭한다. 각 입력 패치에 대하여 입력 영상과 출력 영상간의 우도(likelihood)를 최대화하는 고해상도 패치를 매칭 결과로 선택한다. 따라서, 입력 영상과 무관한 잡영이 발생하는 것을 방지한다. 또한, 영상 관찰 모델과 기존의 패치 키 기반 방법을 MAP의 형태로 결합한 매칭 기준을 함께 제안한다. MAP 매칭 기준은 두 방법의 상보적 특성을 결합하여 잡영을 효과적으로 억제하면서도 영상의 세부 형태를 효과적으로 복원한다. 그 결과 출력 영상의 화질을 개선할 뿐 아니라, 폰트의 종류 및 크기에 대한 적응력도 향상시킨다.

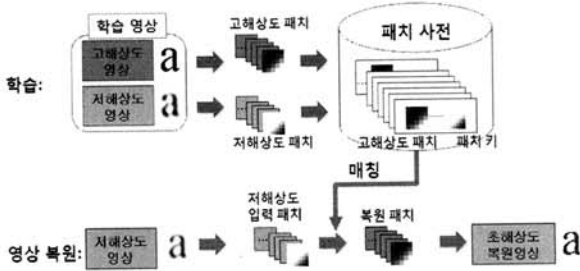
2장에서는 초해상도 영상 복원과 영상 관찰 모델에 대한 기존 연구를 간단히 소개한다. 3장에서는 제안하는 패치 매칭 방법을 설명한다. 4장에서는 실험 결과를 제시하고 마지막 5장에서는 결론을 맺는다.

2. 기존 연구

2.1 예제기반 초해상도 영상복원

다양한 초해상도 영상 복원 방법들 중에서도 EBSR과 베이지안 초해상도 영상 복원은 매우 활발하게 연구되었다 [2-6]. (그림 1)은 EBSR 시스템의 일반적인 구조이다. EBSR 시스템은 고해상도 영상과 저해상도 영상간의 차이를 학습한 후, 이를 이용하여 고해상도 영상의 세부 형태를 복원한다. 따라서, EBSR의 가장 중요한 이슈는 고해상도 영상의 세부 형태 정보를 표현, 저장 및 검색하는 것이다.

영상의 세부 형태는 주로 고주파 성분에 의해 표현된다. Freeman은 이를 이용하여 푸리에 변환을 통해 고해상도 영상의 세부 형태 정보를 추출하였다[4]. 학습 단계에서는 고해상도 영상으로부터 고주파 성분을 추출한 후 이를 작은 패치들로 분할하였다. 그리고, 고주파 패치에 약간의 전처리



(그림 1) 예제기반 초해상도 영상 복원

를 가한 후, 패치 키와 함께 패치 사전에 저장하였다. 패치 키는 중간 주파수 성분으로 구성하였다. 패치 사전은 $\{(k_i, h_i) \mid 1 \leq i \leq N\}$ 와 같이 나타낼 수 있다. 여기에서 h_i 는 i 번째 고주파수 패치이고, k_i 는 h_i 의 키이다. N 은 패치 사전에 포함된 패치들의 수이다.

저해상도 입력 영상 Y 가 주어지면, 이를 입력 패치 y_j ($1 \leq j \leq M$)로 나눈 다음, 입력 패치로부터 키를 추출한다. 여기에서 M 은 입력 패치의 수이다. 그리고, 입력 패치들을 패치 사전의 고주파수 패치들과 매칭한다. 그 후, 각 입력 패치를 그와 매칭된 고주파수 패치들을 결합하여 복원 패치를 생성한다. 입력 패치를 y_i 이라 매칭된 고주파수 패치를 h_i 라고 할 때, 복원 패치 $x_j(h_i)$ 는 식 (1)과 같이 생성된다.

$$x_j(h_i) = F^{-1}(L(y_j) + h_i) \quad (1)$$

$L(y_j)$ 는 y_j 의 저주파수 성분을, F^{-1} 은 역푸리에 변환을 의미한다. 복원 패치가 생성되면 이들을 결합하여 고해상도 출력 영상을 생성한다.

각 입력 패치의 매칭은 입력 패치의 키와 가장 유사한 키를 갖는 고주파 패치를 사전에서 검색함으로써 이루어진다. 그러나, 중간 주파수 성분과 고주파수 패치 간의 상관 관계가 충분히 강하지 않기 때문에 Freeman은 인접한 고주파수 패치간의 상관 관계를 매칭에 반영하였다. 이웃한 패치들의 경계 부분을 겹치게 한 후, 겹치는 영역 내에서의 양쪽 패치간 일치성을 매칭의 목적 함수에 반영하였다. 매칭을 위해서는 두 가지 알고리즘을 제안하였다. 하나는 베이저안 신뢰도 전파(Bayesian belief propagation) 방법이고, 다른 하나는 이를 단순화 한 원패스(one-pass)방법이다.

Freeman이 고주파 성분을 통해 영상의 세부 형태 정보를 나타낸 것에 반해 Chekhlov는 고해상도 영상과 저해상도 영상간의 차 영상을 이용하여 영상의 세부 형태를 반영하였다[7]. 학습에서는 고해상도 영상의 화질을 저해상도 영상으로 저해상도 영상을 생성하였다. 그리고, 저해상도 영상을 고해상도 영상으로부터 뺄으로써 차영상을 생성하였다. 이렇게 생성한 차영상은 고주파수 성분과 마찬가지로 화질 저하 시 소실되는 영상의 세부 형태를 반영한다. 따라서, Chekhlov는 차영상을 패치로 나누어 패치 사전에 저장한 후 이를 이용하여 영상을 복원하였다. 그 외의 알고리즘은 Freeman의 방법과 유사하다.

2.2 영상 관찰 모델과 베이저안 초해상도 영상 복원

영상 관찰 모델(image observation model: IOM)은 베이저안 초해상도 영상 복원을 비롯한 많은 영상 처리 방법에서 사용되었다[2]. 베이저안 초해상도 영상 복원은 입력 영상에 대한 후행 확률을 최대화하는 고해상도 영상을 생성한다[3]. 동일한 장면을 촬영한 여러 장의 저해상도 입력 영상 $Y=(Y^1, Y^2, Y^3, \dots, Y^k)$ 이 주어졌을 때 고해상도 영상 X 의 후행 확률은 $P(X|Y)$ 로 표시된다. 베이저의 정리를 적용하고 X 에 독립인 분모를 제거하면 $P(X|Y)$ 를 최대화 하는 고해상도 영상 X^* 는 식 (2)와 같이 유도된다.

$$\begin{aligned} X^* &= \arg \max_x P(X | Y) \\ &= \arg \max_x \frac{P(Y | X)P(X)}{P(Y)} \\ &= \arg \max_x P(Y | X)P(X) \end{aligned} \quad (2)$$

여기에서 우도(likelihood) $P(Y|X)$ 는 영상 관찰 모델에 의해 유도된다. 널리 사용되는 영상 관찰 모델은 (그림 2)와 같다[2]. 고해상도 영상 X 로부터 저해상도 영상 Y_k 가 관찰되는 과정은 왜곡, 블러링, 다운샘플링 및 가산잡영에 의해 모델링 된다.

여기에서 M^k , B^k 및 D 는 각각 왜곡, 블러링 및 다운샘플링 매트릭스를 나타내며 n^k 는 가우시안 가산 잡영을 나타낸다. 이와 같은 영상 관찰 모델에서 고해상도 영상 X 와 그로부터 관찰된 저해상도 영상 Y^k 의 관계는 식 (3)와 같다.

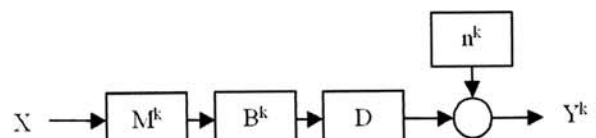
$$Y^k = DB^k M^k X + n^k \quad (3)$$

식 (3)는 화질 저하 매트릭스 $W^k = DB^k M^k$ 를 도입함으로써 식 (4)과 같이 단순화 될 수 있다.

$$Y^k = W^k X + n^k \quad (4)$$

매트릭스 M^k , B^k 및 D 는 평행이동 벡터, 회전각, 다운샘플링 비율 등의 파라미터에 의해 결정되며 이러한 파라미터들은 다양한 영상 등록 알고리즘에 의해 추정할 수 있다[8]. 이와 같은 영상 관찰 모델로부터 우도 $P(Y|X)$ 는 식 (5)와 같이 정의된다. 여기에서 σ^2 는 가우시안 잡영의 분산이다.

$$\begin{aligned} P(Y | X) &= \prod_k P(Y^k | X) \\ P(Y^k | X) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\|Y^k - W^k X\|^2}{2\sigma^2}\right] \end{aligned} \quad (5)$$



(그림 2) 영상 관찰 모델

대부분의 경우 (5)를 최대화 하는 X 를 찾는 문제는 충분하지 않은 입력 영상으로 인해 정보가 부족한(ill-posed) 문제가 된다. 베이지안 초해상도 영상 복원에서는 선행 확률을 적절히 정의함으로써 문제를 보완하였다. Huber의 에지 보존적 평활화 선행 함수와 Donaldson의 바이모달(bimodal) 선행 함수는 다양한 시스템에 널리 사용되고 있다[9, 10].

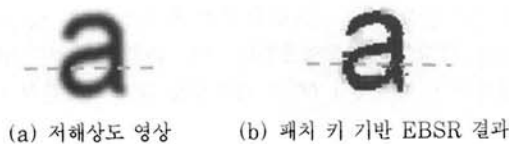
Park은 베이지안 초해상도 영상 복원과 EBSR을 결합한 영상 복원 시스템을 제안하였다[5]. 먼저 EBSR을 이용해 고해상도 참조 영상 G 을 생성한 후, 그 결과를 베이지안 초해상도 영상 복원의 선행 확률에 반영하였다. 선행 확률은 최종 복원 영상 X 와 G 의 유사도에 의해 정의되었다. EBSR 복원 영상에서 유도된 선행 확률을 통해 패치 사전의 정보를 활용하는 동시에 식 (5)의 우도를 통해 입력 영상과의 상관 관계를 반영하였기 때문에 두 방법의 장점을 모두 갖는다.

3. 영상 관찰 모델을 이용한 초해상도 텍스트 영상 복원

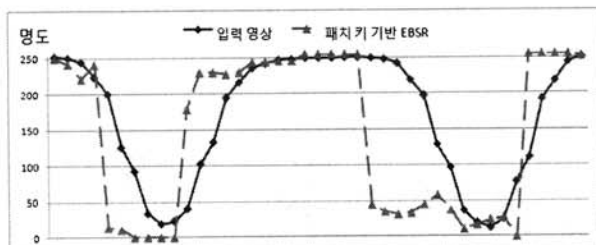
3.1 영상 관찰 모델을 이용한 패치 매칭 기준

1장에서 기술한 잡영 문제는 입출력 영상의 상관 관계를 반영함으로써 극복될 수 있다. (그림 3)-(b)는 (그림 3)-(a)에 대한 EBSR 결과이다. (그림 3)-(b)에는 (그림 3)-(a)에 비해 세부 형태가 많이 복원되었으나, 잘못 매칭된 패치들로 인해 잡영도 포함되었다. (그림 4)는 (그림 3)을 점선을 따라 절단한 단면이다. 가로축은 x 좌표를, 세로축은 각 화소의 명도를 나타낸다. 잡영이 발생한 좌표에서 입력 영상과 출력 영상의 명도가 크게 다른 것을 볼 수 있다. 이 같은 잡영은 입출력 영상간의 상관 관계를 패치 매칭 과정에 반영함으로써 억제할 수 있다.

본 연구에서는 입출력 영상간 상관 관계를 패치 수준의 영상 관찰 모델(IOM)을 이용하여 패치 매칭 단계에 반영하였다. EBSR 시스템은 입력 패치를 그와 매칭된 고해상도



(그림 3) 패치 키 기반 EBSR 결과



(그림 4) 그림 3의 단면

도 패치와 결합하여 고해상도 복원 패치를 생성한다. 이때, 입력 패치 y_j 에 대한 고해상도 패치 h_i 의 우도는 2,2절에서 기술한 영상 관찰 모델에 의해 식 (6)와 같이 정의할 수 있다.

$$P(y_j | h_i) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{\|y_j - Wx_j(h_i)\|^2}{2\sigma^2}\right] \quad (6)$$

여기에서 $x_j(h_i)$ 는 y_j 와 고해상도 패치 h_i 로부터 식 (1)을 통해 생성한 복원 패치를 나타내며, W 는 식 (4)에서 소개한 화질 저하 매트릭스를 나타낸다.

최대 우도(maximum likelihood) 방식을 적용하면 각 입력 패치 y_j 는 식 (6)를 최대화하는 고해상도 패치와 매칭된다. 이러한 고해상도 패치 h_j^* 는 식 (7)과 같이 유도할 수 있다.

$$\begin{aligned} h_j^* &= \arg \max_h P(y_j | h_i) \\ &= \arg \max_h \log P(y_j | h_i) \\ &= \arg \min_h \|y_j - Wx_j(h_i)\|^2 \end{aligned} \quad (7)$$

식 (7)에 의해 각 입력 패치에 고해상도 패치가 매칭되면 [4]과 동일한 방법으로 식 (1)을 통해 복원 패치를 생성하고 이들을 결합함으로써 고해상도 영상을 복원한다. (그림 5)는 제안하는 방법과 기존의 패치 키 기반 방법을 (그림 3)-(a)에 적용한 결과이다. 두 방법간 차이를 강조하기 위하여 (그림 3)-(b)보다 더 큰 패치 사전을 이용하였다. 패치 키를 이용한 방법이 (그림 5)-(b)와 같이 많은 잡영을 발생시킨 반면, IOM을 이용한 방법은 (그림 5)-(a)와 같이 잡영을 성공적으로 억제하였다.

그러나, (그림 5)-(a)에서는 에지 등 영상의 세부 형태가 잘 복원되지 않았다. 그 이유는 베이지안 초해상도 영상 복원에 관한 여러 문헌에서 지적된 바와 같이 식 (6)를 최대화 하는 문제가 정보가 부족한(ill-posed) 문제이기 때문이다 [2, 3]. 이를 극복하기 위하여 본 연구에서도 베이지안 초해상도 영상 복원에서와 마찬가지로 선행 함수를 통해 식 (6)의 정보 부족을 보완하였다.

패치 키 기반 방법에 의해 복원된 (그림 5)-(b)에는 많은 잡영이 발생되었으나, 에지 성분이 성공적으로 복원 것을



(a) IOM을 이용한 EBSR (b) 패치 키 기반 EBSR

(그림 5) IOM을 이용한 EBSR과 패치 키 기반 EBSR 결과

볼 수 있다. 이는 패치 키 기반 패치 매칭이 세부 형태 복원에 도움이 된다는 것을 보여준다. 따라서, 본 연구에서는 패치 키간의 차이를 이용하여 선행 확률을 정의하였다. 제안하는 선행 확률은 식 (8)과 같이 패치 키간 차이에 대한 가우시안 함수로 정의된다.

$$P(h_i) = \frac{1}{\sqrt{2\pi\lambda}} \exp\left[-\frac{\|k_j - k_i\|^2}{2\lambda^2}\right] \quad (8)$$

여기에서 k_j 와 k_i 는 각각 입력 패치 y_j 와 고해상도 패치 h_i 의 키를 나타낸다.

식 (6)과 식 (8)으로부터 y_j 에 대한 후행 확률을 최대화 (maximum a posterior: MAP)하는 고해상도 패치 h_j^* 는 식 (9)와 같이 유도된다.

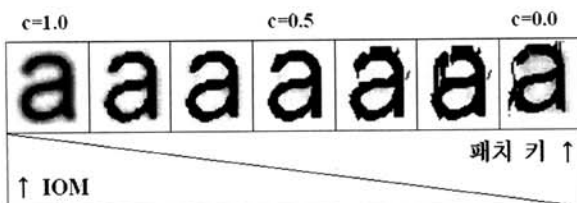
$$\begin{aligned} h_j^* &= \arg \max_h P(h_i | y_j) \\ &= \arg \max_h P(y_j | h_i) P(h_i) \\ &= \arg \max_h [\log P(y_j | h_i) + \log P(h_i)] \\ &= \arg \min_h \left[\frac{1}{2\sigma^2} \|y_j - Wx_j(h_i)\|^2 + \frac{1}{2\lambda^2} \|k_j - k_i\|^2 \right] \end{aligned} \quad (9)$$

여기에서 $c = \lambda^2 / (\sigma^2 + \lambda^2)$ 로 정의할 경우 식 (9)는 식 (10)과 같이 정리할 수 있다.

$$h_j^* = \arg \min_h \left[c \|y_j - Wx_j(h_i)\|^2 + (1-c) \|k_j - k_i\|^2 \right] \quad (10)$$

MAP 매칭 기준 (10)은 IOM기반 매칭 기준과 패치 키 기반 매칭 기준의 가중 평균으로 구성된다. 전자는 입출력 영상간의 상관 관계를 반영하며 후자는 전자에서 부족한 정보를 보충함으로써 텍스트 영상의 세부 형태가 더 잘 복원되게 한다. 상수 c ($0 \leq c \leq 1$)는 두 매칭 기준의 비중을 결정하는 가중치의 역할을 한다. (그림 6)은 가중치 c 를 변화시키며 MAP 매칭 기준을 이용한 영상 복원 결과이다.

c 가 1일 경우 복원 결과는 (그림 5)-(a)과 동일하다. c 가 감소함에 따라 복원 결과는 (그림 5)-(b)에 가까워진다. 위의 예에서는 $c=0.5$ 일 때 가장 좋은 결과를 출력하였다. 이에 해당하는 가운데 영상은 잡영이 억제되었으면서도 예지 성분이 효과적으로 복원되었다.



(그림 6) 다양한 가중치에 따른 MAP 기준에 의한 복원 결과

3.2 패치 매칭 알고리즘

(그림 6)에서 볼 수 있듯이, 제안하는 MAP 매칭 기준을 사용할 경우 안정적으로 패치를 매칭하여 우수한 복원 성능을 얻을 수 있다. 그러나, 이는 많은 연산을 필요로 하기 때문에 영상 복원 속도를 저하시킨다. 사전 실험 결과, 제안하는 방법에 의하여 패치를 비교할 경우 패치 키를 사용한 경우보다 약 5배 느렸다. 그러나, 제안하는 방법을 사용할 경우 복잡도가 높은 신뢰도 전과 방법을 사용하지 않아도 우수한 매칭 결과를 얻을 수 있기 때문에, 전체적인 복원 속도는 심하게 저하되지 않는다. 제안하는 매칭 기준을 이용할 경우 패치 간 비교만으로도 안정적으로 패치를 매칭할 수 있기 때문이다. 그러나, 매칭 속도의 개선은 역시 필요하다.

본 연구에서는 빠른 속도로 패치를 매칭하기 위해 두 단계로 이루어진 매칭 알고리즘을 사용하였다. 먼저, 각 입력 패치에 대하여 패치 사전의 고해상도 패치들로부터 패치 키 비교를 통해 소수의 매칭 후보들을 선택하였다. 그리고, 선택된 후보들로부터 제안하는 매칭 기준을 이용해 최적의 매칭 결과를 선택하였다. 대용량의 패치 사전으로부터 소수의 후보 패치를 선택하기 위해서는 [11]에 소개된 다단계 후보 축소 방법을 사용하였다. 이와 같은 방법에 의해 구성된 EBSR 시스템은 패치 키 기반 시스템과 유사한 속도를 나타내었다.

4. 실험

제안하는 방법의 성능을 평가하기 위하여 두 가지 실험을 수행하였다. 첫 번째 실험은 복원 영상의 화질 평가를 위한 실험이고, 두 번째 실험은 제안하는 방법을 텍스트 영상 인식 시스템의 전처리 단계에 적용하였을 때 인식 성능에 미치는 영향을 평가하기 위한 실험이다.

4.1 실험 데이터 및 환경

실험에는 두 종류의 문자 영상 데이터베이스를 사용하였다. 첫 번째 데이터는 ETRI 인쇄 문자 영상 데이터이고[13], 두 번째 데이터는 윈도우 폰트로부터 수집한 데이터이다. ETRI 데이터는 공개된 데이터이나, 다양한 폰트를 포함하지 않는다. 따라서, 이를 보완하기 위하여 다양한 종류와 크기의 폰트를 포함한 영상 데이터를 추가로 수집하였다. 본 연구에서 수집한 데이터는 Arial, Century, Courier New, Tahoma, Times New Roman, Verdana, Lucida Sans, 돋움, 굴림, 궁서, 고딕 등 11가지 폰트를 포함하고 있으며, 각 폰트에 대하여 9pt에서 26pt까지 18가지 다른 크기로 출력된 영상들로 구성된다. 각 데이터베이스에 포함된 샘플들의 수는 <표 3>과 같다.

패치 사전은 폰트 데이터만으로 학습하였고, 인식기 학습에는 두 가지 데이터를 모두 사용하였다. 학습 데이터와 테스트 데이터는 폰트의 크기를 기준으로 나누었기 때문에, 서로 다른 크기의 문자 영상들로 구성되었다. 인식기는 '영 어 대소문자와 숫자 등 62개의 클래스를 대상으로 하였다.

〈표 3〉 실험에 사용한 데이터

데이터 베이스	폰트 종류	학습 영상		테스트 영상	
		클래스당 샘플수	전체 샘플수	클래스당 샘플수	전체 샘플수
ETRI	2~3	166	10,292	84	5,208
폰트 데이터	11	180	11,160	90	5,580

실험은 인텔 Q6600 2.4GHz와 4GB의 메모리를 갖는 PC에서 수행하였다.

4.2 영상 복원 실험

다양한 저해상도 영상에 제안하는 방법을 이용하여 고해상도 영상을 복원한 후, 기존의 방법에 의한 결과와 비교하였다. 본 실험에서는 제안하는 방법을 최소 거리 이웃(nearest neighbor) 및 3차 운형(cubic-spline) 등 두 가지 보간 방법 [14]과 Freeman의 영상 복원방법[4], 박장균 의 방법[5] 등 두 가지 초해상도 영상 복원 방법과 비교하였다. 보간 방법은 저해상도 영상으로부터 고해상도 영상을 생성하는 가장 기본적인 방법으로, 영상 복원 시스템의 성능 평가 시 비교 대상으로 많이 사용된다. Freeman의 방법은 EBSR 방법 중 가장 대표적인 것 뿐 아니라 제안하는 방법의 기반이 된다. 박장균 의 방법은 패치 사전의 정보를 사용하면서도 영상 관찰 모델을 통해 입력 영상과의 상관 관계를 활용한다는 점에서 제안하는 방법과 유사점이 있다. 그러나, 박장균 의 방법이 EBSR의 결과를 베이지안 초해상도 영상 복원에서 사용하는 반면, 제안하는 방법은 베이지안 초해상도 영상 복원에서 많이 사용되는 영상 관찰 모델을 EBSR의 패치 매칭 단계에 사용했다는 점에서 크게 다르다. 또한, 박장균의 방법은 영상 관찰 모델을 이용하기는 하지만 EBSR 단계에서는 패치 키를 이용하여 패치를 매칭하기 때문에 1장에서 설명한 문제들에 대한 완전한 해결책은 제시하지 않는다.

먼저 고해상도 영상에 가우시안 블러링과 다운샘플링을 통해 저해상도 영상을 생성하였다. 가우시안 블러링의 표준편차는 문자 영상 높이의 4%로 하였으며, 다운샘플링 비율은 1/2로 하였다.

〈표 4〉-(a)와 (b)는 고해상도 원본 영상 및 그로부터 생성한 저해상도 영상의 예이다. 그리고, 기존의 방법과 제안하는 방법을 이용해 저해상도 영상으로부터 고해상도 영상을 복원하였다. 〈표 4〉-(c)~(g)는 그 결과이다. (c)와 (d)는 각각 최소 거리 이웃보간 및 3차 운형 보간의 결과이다[14]. (e)는 기존 패치 키 기반 방법의 결과이다[4]. (f)는 (e)을 선행 확률에 반영한 베이지안 초해상도 영상 복원결과로 (e)에 비해 잡영이 다소 감소하였다[5]. (g)와 (h)는 제안하는 방법의 결과이다. (g)는 식 (7)의 IOM만을 이용한 결과이며, (h)는 식 (10)의 MAP 매칭 기준(c=0.15)을 사용한 결과이다. (e)에서는 에지 성분이 잘 복원되었으나, 잡영이 많이 발생하였다. (g)는 잡영이 효과적으로 억제되었으나, 에지 성분을 잘 복원하지 못하였다. (h)는 두 방법의 상보적 특성을 결합하여 가장 좋은 결과를 나타내었다.

〈표 4〉 영상 복원 결과

-	영상 복원 결과
(a) 고해상도	
(b) 저해상도	
(c) Nearest-neighbor [14]	
(d) Cubic-spline [14]	
(e) 패치 키[4]	
(f) 박장균[5]	
(g) IOM	
(h) MAP	

정량적 비교를 위하여 각 복원 영상과 고해상도 원본 영상간의 차이를 측정하였다. 비교를 위한 척도로는NRMSE (normalized root mean square error)을 사용하였다 [15]. 두 영상 간의 NRMSE는 식 (11)같이 정의된다. A 와 B는 두 개의 MxN영상을 나타내며, A_{ij}와 B_{ij}는 각각 A와 B의 i번째 행, j번째 열의 화소를 의미한다.

다양한 종류의 패치 사전에 대한 성능을 평가하기 위하여 네 종류의 패치 사전을 이용하여 실험하였다. 첫 번째 실험에서는 단일 폰트, 단일 크기의 영상으로 학습된 사전들을 이용하였다. 영상 복원에는 입력 영상과 동일한 종류 및 크기의 폰트로 학습된 사전을 사용하였다. 두 번째 실험에서는 다중 폰트, 단일 크기의 영상으로 학습된 사전들을 이용하였다. 영상 복원에는 입력 영상과 동일한 크기의 영상으로 학습된 사전을 이용하였다. 세 번째 실험에서는 반대로 단일 폰트, 다중 크기의 영상으로 학습된 사전을 이용하였다. 영상 복원에는 입력 영상과 동일한 폰트로 학습된 사전을 이용하였다. 마지막으로 네 번째 실험에서는 모든 폰트와 크기가 포함된 한 개의 대용량 사전을 이용하였다. 실험 결과는 〈표 5〉와 같다.

EBSR의 복원 성능은 보간법보다 우수한 것으로 나타났다. 패치 키 기반 방법에서는 패치 사전의 크기가 증가함에 따라 NRMSE가 증가하였다. 이는 패치 매칭의 오류가 패치 사전의 크기와 함께 증가했음을 보여준다. 박장균 의 방법은 패치 키 기반 방법에 비해 다소 나은 성능을 보였다. IOM을 이용한 방법에서는 기존 방법들에 비해 NRMSE가 더욱 감소하였을 뿐 아니라, 패치 사전의 크기가 증가하

〈표 5〉 복원 영상과 원본 영상간 NRMSE

	단일 크기		다중 크기 (9~26pt)	
	단일 폰트	다중폰트 (11종)	단일 폰트	다중폰트 (11종)
Nearest-neighbor	0.334			
Cubic-spline	0.328			
패치 키	0.241	0.287	0.301	0.307
박장균	0.277	0.274	0.294	0.256
IOM	0.248	0.243	0.245	0.238
MAP	0.226 (c=0.15)	0.138 (c=0.30)	0.243 (c=0.15)	0.157 (c=0.10)

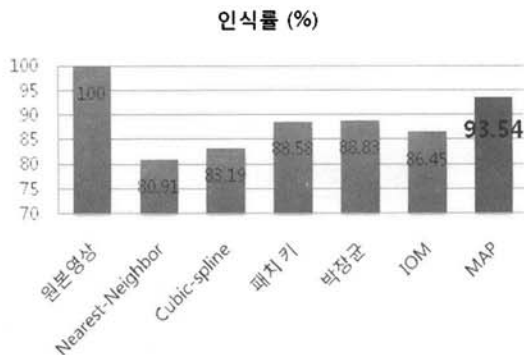
라도 NRMSE가별로 증가하지 않았다. 이는 패치 사전이 매우 클 때에도 패치들을 안정적으로 매칭했다는 것을 보여준다. MAP 매칭 기준의 경우 IOM만을 이용한 경우보다 더욱 우수한 결과를 나타내었다. 특히, 패치 사전이 다양한 폰트로 학습된 경우에 단일 폰트로 학습된 경우보다 더욱 우수한 성능을 나타내었는데, 이는 MAP 매칭 기준이 다양한 폰트의 패치들을 영상 복원에 효과적으로 사용하였음을 의미한다.

4.3 인식 성능

제안 하는 방법을 인식 시스템에 적용하였을 때의 실제적 효과를 측정하기 위하여 문자 인식을 구현한 후 각 복원 영상들을 인식하여 그 인식률을 측정하였다. 문자 인식기는 최소 거리 구분 방법(minimum distance classifier)을 이용해 구현하였다[16]. 특징 추출 단계에서는 성능이 우수하다고 알려진 기울기(gradient) 특징을 이용하였으며, NCGF(normalization-cooperated gradient feature) 알고리즘을 이용해 추출하였다[17]. 특징 벡터는 8가지 방향에 대하여 8x8 그물망을 통해 추출되었으며 그 결과, 8x8x8 = 512차원으로 구성되었다. 또한, 구분력을 향상하기 위하여 LVQ(learning vector quantization)를 적용하였다[18].

각 방법에 의해 복원된 영상에 대하여 테스트한 인식률은 <표 6>과 같다. 고해상도 원본 영상에 대한 인식률은 100%로 나타났다. 그러나, 복원 영상에 대한 인식률은 그 보다

〈표 6〉 각 복원 방법에 대한 인식성능



훨씬 낮았다. 초해상도 영상 복원 방법들은 이번에도 보간법에 의한 영상 복원보다 우수한 성능을 나타내었다. IOM만을 이용한 방법은 기존의 패치 키 기반 방법과 비교할 때 화질 면에서는 다른 특성을 보였으나, 인식 성능은 유사하였다. MAP 매칭 기준의 경우 93.54%의 인식률을 나타내어 가장 우수한 성능을 보였다. 88.58%의 인식률을 보인 패치 키 기반 방법과 비교할 때 43.4%의 오류감소율을 나타내었다.

5. 결론

초해상도 영상 복원은 다른 영상 복원 방법에 비하여 여러 가지 장점을 갖는다. 그러나, 폰트의 종류 및 크기가 학습 영상과 다른 텍스트 영상에 적용하였을 때에는 많은 잡영을 발생시킨다. 이러한 문제의 원인은 입력 패치들이 패치 사전 내 고해상도 패치들과 잘 못 매칭될 수 있기 때문이다.

본 논문에서는 텍스트 영상의 EBSR을 위한 새로운 매칭 기준을 제안하였다. 제안하는 방법은 패치 수준의 영상 관찰 모델을 이용하여 입출력 영상의 상관 관계를 반영함으로써 효과적으로 잡영을 억제한다. 그 결과, 잡영의 발생을 억제할 뿐 아니라, 대용량 패치 사전을 적용 가능하게 함으로써 폰트의 종류 및 크기에 대한 적응력을 향상시킨다. 또한, 영상 관찰 모델과 패치 키 기반 방법을 결합한 MAP 매칭 기준을 제안하였다. MAP 매칭 기준은 두 방법의 상보적 특성을 결합함으로써 효과적으로 잡영을 억제하면서도 영상의 세부 형태를 잘 복원한다.

실험을 통해 제안하는 방법의 영상 복원 성능을 기존의 방법과 비교하였다. 제안하는 방법은 화질 복원 성능과 인식 성능 모두에서 기존 방법보다 우수한 성능을 나타내었다. 특히, 인식 실험에서는 93.54%의 인식률을 보여 88.58%의 인식률을 보인 기존의 방법과 비교할 때 43.4%의 오류감소율을 보였다.

참고 문헌

- [1] 류상진, 김인중, “저화질 영상 인식을 위한 화질 저하 모델 기반 다중 인식기 결합”, 정보처리학회논문지B, 제17-B권, 제3호, pp.233-238, 2010.
- [2] S. C. Park, M. K. Park, and M. G. Kang, “Super-resolution image reconstruction: a technical overview,” IEEE Signal Processing Magazine, Vol.20, No.3, pp.21-36, 2003.
- [3] M. E. Tipping and C. M. Bishop, “Bayesian image super-resolution,” Advances in Neural Information Processing Systems 15, MIT Press, Cambridge, MA, pp.1279-1286, 2003.
- [4] W. T. Freeman, T. R. Jones, and E. C Pasztor, “Example-Based Super-Resolution,” IEEE Computer Graphics and Applications, Vol.22, No.2, pp.56-65, 2002.
- [5] J.Park, Y.Kwon, J.Kim, “An Example-Based Prior Model for Text Image Super-resolution,” Proc. of ICDAR2005, Seoul,

Korea, pp.374-378, 2005.

[6] L. C. Pickup, D. P. Capel, S. J. Roberts, and A. Zisserman, "Bayesian image super-resolution, continued," Advances in Neural Information Processing Systems 19, pp.1089-1096, Cambridge, Mass, USA, December, 2006.

[7] D. Chekhlov, "Super-Resolution of Images," Ph.D. Thesis, Bristol, pp.1-12, 2005.

[8] L.G. Brown, "A survey of image registration techniques," ACM. Computing Surveys, Vol.24, pp.326-376, 1992.

[9] D. P. Capel, "Image Mosaicing and Super-resolution," Ph.D. thesis, University of Oxford, 2001.

[10] K. Donaldson and G. K. Myers, "Bayesian Super-Resolution of Text in Video with a Text-Specific Bimodal Prior," IJDAR, Vol.7, No.2, pp.1433-2833, 2005.

[11] <http://people.csail.mit.edu/hasinoff/320/sliding-notes.pdf>

[12] 박규로, 김인중, "단계적 후보 축소에 의한 예제기반 초해상도 영상복원을 위한 고속 패치 검색", 정보과학회논문지, Vol.37, No.4, pp.264-272, 2010.

[13] <http://ai.kaist.ac.kr/Resource/dbase/Image%20Database.htm>

[14] <http://en.wikipedia.org/wiki/Interpolation>

[15] <http://en.wikipedia.org/wiki/RMSE>

[16] R.O. Duda, P.E. Hart and D.G. Stork, Pattern Classification 2nd ed., pp.36-45, Wiley-Interscience, 2001.

[17] C. L. Liu, "Normalization-Cooperated Gradient Feature Extraction for Handwritten Character Recognition," IEEE TPAMI, Vol.29, No.8, pp.1465-1469. 2007.

[18] C. L. Liu, I. J. Kim, and J. H. Kim, "High Accuracy Handwritten Chinese Character Recognition by Improved Feature Matching Method," Proc. 4th ICDAR, Ulm, Germany, pp.1033-1037, 1997.



박규로

e-mail : outzzang@gmail.com

2008년 2월 한동대학교(학사)

2008년 3월~현 재 한동대학교 정보통신공학과 석사과정

관심분야: 영상처리, 패턴인식, 초해상도 영상복원, 저화질 영상인식



김인중

e-mail : ijkim@handong.edu

1994년 2월 KAIST 전산학과(학사)

1995년 8월 KAIST 전산학과(석사)

2001년 2월 KAIST 전산학과(박사)

2001년 2월~2006년 2월 (주)인지소프트 객원연구원

2006년 3월~현 재 한동대학교 교수

관심분야: 패턴인식, 영상처리, 인공지능, 모바일 프로그래밍, 증강현실