

Noise reduction algorithm for an image using nonparametric Bayesian method

Ho-young Woo^a · Yeong-hwa Kim^{a,1}

^aDepartment of Applied Statistics, Chung-Ang University

(Received April 9, 2018; Revised June 14, 2018; Accepted August 1, 2018)

Abstract

Noise reduction processes that reduce or eliminate noise (caused by a variety of reasons) in noise contaminated image is an important theme in image processing fields. Many studies are being conducted on noise removal processes due to the importance of distinguishing between noise added to a pure image and the unique characteristics of original images. Adaptive filter and sigma filter are typical noise reduction filters used to reduce or eliminate noise; however, their effectiveness is affected by accurate noise estimation. This study generates a distribution of noise contaminating image based on a Dirichlet normal mixture model and presents a Bayesian approach to distinguish the characteristics of an image against the noise. In particular, to distinguish the distribution of noise from the distribution of characteristics, we suggest algorithms to develop a Bayesian inference and remove noise included in an image.

Keywords: adaptive filter, Bayesian statistics, Dirichlet normal mixture model, image processing, noise reduction

1. 서론

최근 들어 디지털 카메라, 스마트폰 카메라, CCTV, 블랙박스 등과 같은 수 많은 영상기기는 실생활에서 빼놓을 수 없는 중요한 도구로 자리 잡고 있다. 이러한 영상기기를 디지털 영상기기라고 하며 단순한 상(像)을 의미했던 영상의 의미는 시각적인 정보에 대한 획득과 표현을 총칭하는 의미로 확장되었으며 과학기술의 발달로 인해 디지털의 영역에서 다루고 있다. 특히 디지털 영상을 다루는 것을 영상처리(image processing)라고 하며 영상처리는 특정 영상을 사용자의 목적에 부합하도록 개선하거나 변환하는 목적 이외에도 원래의 순수한 영상을 최대한 본래의 모습 그대로 구현하고자 하는 것도 영상처리의 중요한 목적이 되고 있다. 그러나 원영상을 기록하거나 다시 나타내는 전반적인 과정에서 다양한 원인에 의하여 잡음(noise)이 추가될 수 있으며, 잡음은 영상에서 구현하고자 하는 모서리(edge)나 세밀한 부분(small detail) 등과 같은 중요 성분인 특징(feature)을 구별하기 어렵게 한다. 기술의 발전에도 불구하고 영상에는 이러한 잡음이 추가될 수 있으며 추가된 잡음은 영상의 질을 떨어뜨리는 중요한 요인이 된다. 영상처리 분야에서 잡음 제거(noise reduction)는 잡음의 특성과 잡음의 분포에 따라 활발한 연

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2017R1D1A1B03031725).

¹Corresponding author: Department of Applied Statistics, Chung-Ang University, 84, Heukseok-ro, Dongjak-gu, Seoul 06974, Korea. E-mail: gogators@cau.ac.kr

구가 진행되고 있으며 잡음의 분포에 대한 가정에 따라 가우시안(Gaussian) 잡음, 균일 잡음 등의 잡음 모형이 있다. 일반적으로 가우시안 잡음 제거에 관한 연구는 잡음의 분포가 정규분포를 따른다는 가정 하에서 잡음의 분산을 추정하고 이를 이용하여 잡음을 제거하는 과정이 포함된다 (González와 Woods, 2008). Kim과 Nam (2011)은 영상의 방향성에 대한 분산의 동일성 검정을 활용하여 영상에 포함된 잡음을 제거하는 알고리즘을 제시하였으며, Kim (2012)은 블록방법에 근거하여 적응적 잡음제거 알고리즘을 제시하였다. Beak 등 (2015)는 저 계수 행렬 근사(low rank approximation)를 통해 잡음 제거 알고리즘의 단점을 보완하는 알고리즘을 제안하였다. 그러나 분포에 대한 가정이 어긋날 경우 잡음 추정 에 있어서 문제가 발생할 수 있으므로 분포에 대한 가정을 완화하는 것이 때로는 추론의 효율을 높이고 잘못된 추론의 결과를 도출할 가능성을 줄일 수 있다. 이와 같은 방법을 비모수적 방법이라고 하며 이는 분포에 대해 특정한 형태를 가정하지 않는 통계적 방법을 말한다. 비모수적 방법은 가정이 위배되더라도 추론의 오류를 줄이며 주어진 자료가 척도불변의 성질을 가지지 않을 때에도 사용 할 수 있는 장점을 가진다 (Song 등, 2003). 그러나 통계량의 분포가 복잡하고 접근 이론을 기반으로 추론을 전개하므로 표본의 크기가 작은 경우에는 사용하기 어렵다.

비모수 베이지안 방법은 디리클레 과정(Dirichlet process; DP)을 사전정보로 활용하여 통계적 모형을 구축한다. 기존의 베이지안 방법은 추론의 대상인 모수를 확률변수로 간주하지만 비모수 베이지안 방법은 사전분포를 확률과정으로 가정하며, DP는 무한한 모수 공간상의 분할(partition)된 집합들에 대한 임의 분포들의 결합 분포가 디리클레 분포를 따른다는 것이다 (Müller와 Quintana, 2004). 만약 DP가 주어졌다면 관측 값들의 분포도 DP를 따르게 된다. 관측 값들이 주어졌을 때 추정하고자 하는 임의의 분포는 관측 값들이 나타날 가능성을 토대로 추정된다. DP는 분포에 대한 분포로 생각할 수 있으며 무한하게 혼합된 분포에 대해서도 접근할 수 있게 하여 이로 인해 임의의 분포에 대한 추정도 가능하게 한다. 비모수 베이지안 방법은 분포 가정에 대한 의존도를 줄이고 다양한 통계 모형에 대해 로버스트한 결과를 제공하며 (Ferguson, 1973), 소표본하에서도 Markov Chain Monte Carlo (MCMC) 방법을 통하여 대표본처럼 추론을 용이하게 한다. 실질적으로 영상에 포함되는 잡음은 분포 가정이 어려워 획득된 단 하나의 영상을 토대로 잡음에 대한 추론을 진행하기 때문에 대표본 가정을 할 수 없다. 비모수적 베이지안 방법은 임의 잡음에 대한 분포 가정을 필요치 않으며 단일 영상을 토대로 잡음에 대한 추론이 가능하므로 기존의 방법에 새로운 대안으로 생각할 수 있다. 본 연구에서는 영상의 분포를 추정하는 비모수 베이지안 방법을 사용하였으며 기존의 적응적 필터에 이러한 비모수 베이지안 방법을 적용하여 영상에 포함된 잡음을 제거하는 방안을 제시하자 한다. 본 논문의 구성은 다음과 같다. 제2장에서는 본 논문에서 다루는 영상처리의 여러 개념과 비모수 베이지안 방법에 대해 설명한다. 제3장에서는 영상의 잡음을 제거하는 방법으로 기존의 적응적 필터에 비모수 베이지안 추론을 사용한 적응적 필터를 제안하고 모의 실험을 통하여 기존에 사용되는 방법들과 비교하였다. 제4장에서는 본 논문을 종합적으로 정리하고 요약하여 논문을 마무리하였다.

2. 영상처리와 비모수 베이지안 추론

2.1. 영상처리

영상(image)이란 시각적인 측면에서 정보를 기록하거나 표현하는 방법이며 일반적으로 동영상보다는 정지 영상인 그림이나 사진을 뜻한다. 컴퓨터 기술과 전자 기술의 비약적인 발전에 따라 영상도 디지털 영역에서 다뤄지고 있으며 이러한 디지털 영상을 다루는 분야를 디지털 영상처리라고 한다 (González와 Woods, 2008). 넓은 의미로서 영상처리는 컴퓨터와 전자기기를 이용하여 영상을 제작, 처리, 해석, 인식하는 과정으로서 영상과 연관된 모든 분야를 말한다. 영상처리 분야 중에서 잡음 제거는 잡음이 포함된 영상에 대하여 사전지식과 다양한 방법을 이용하여 영상에 포함된 잡음을 제거하는 것을 의미하며,

오염된 영상은 원 영상과 잡음이 혼재되어 나타나며 원 영상의 특징과 잡음을 추정하는 것이 중요하다. 오염된 영상은 다음과 같이 나타낼 수 있다.

$$d(x, y) = f(x, y) + n(x, y),$$

여기서 d 는 잡음과 원영상이 혼재되어 나타나는 잡음 영상, f 는 원 영상, n 은 잡음을 나타낸다. 잡음은 영상에 임의의 값이 첨가되는 것을 말하며 정확한 원인을 알 수 없이 임의적으로 발생한다. 잡음은 영상을 획득하거나 처리하는 과정에서 모두 발생할 수 있으며 잡음을 다루는 것은 영상처리에 있어서 중요한 부분이라 할 수 있다.

영상처리 분야에서 특징과 잡음을 구별하기 위해 사용하는 통계량은 통계학에서 분산의 추정치인 표본 분산 혹은 표준편차의 추정치인 표본 표준편차이다. 영상의 잡음 수준이 높고 낮음을 표본 분산 또는 표본 표준편차를 사용하여 확인한다. 그러나 분산의 추정치만을 사용하는 것은 잡음을 효과적으로 제거하는데 한계가 존재한다. 동일한 표본 표준편차를 갖는다 하더라도 어떤 영상은 특징, 어떤 영상은 잡음일 수 있다. 영상을 처리하는 절차 또는 장치를 필터(filter)라고 하는데 여기서 필터는 공간 필터(spatial filter)를 말하며 공간 필터링은 영상에 있는 공간 주파수 대역을 제거하거나 강조하는 필터 처리를 뜻한다. 이러한 공간 필터는 크게 컨벌루션과 같이 선형 변환 과정을 통하는 선형 공간 필터와 선형적인 변환으로 계산할 수 없는 비선형 연산을 기반으로 하는 비선형 공간 필터로 나눌 수 있다. 선형 공간 필터를 이용한 영상처리는 2차원의 선형 변환 과정인 2차원 컨벌루션(convolution)을 수행하며 $m \times n$ 크기의 필터 마스크 h 를 이용한 공간 필터는 다음과 같이 나타낼 수 있다.

$$g(x, y) = \sum_{i=-s}^s \sum_{j=-t}^t h(i, j)f(x + i, y + j),$$

여기서 $s = (m - 1)/2$ 이고, $t = (n - 1)/2$ 이다. 선형 공간 필터에서 사용하는 필터 마스크 또는 회선 마스크는 홀수 크기를 가진 정방형 마스크가 사용되며 사용되는 필터 마스크의 계수 값들에 따라 영상을 강조하거나 평활화시키므로 마스크의 계수 값에 따라 공간 필터의 특징을 결정짓는다. 필터 마스크의 계수 값의 합은 1되어야 하므로 계수 값은 가중치로 생각할 수 있다. 공간 필터에는 저주파 성분을 남기고 고주파 성분을 제거하는 저주파 통과 필터, 고주파 성분을 남기고 저주파 성분을 제거하는 고주파 통과 필터 등이 있다. 저주파 통과 필터는 신호 성분 중 저주파 성분은 통과시키고 고주파 성분은 차단하므로 주로 잡음을 제거하거나 열화된(blurring) 영상을 얻을 때 사용하는 필터이다. 저주파 통과 필터는 마스크의 모든 계수가 양수이고 전체 합이 1인 마스크가 사용된다. 저주파 통과 필터를 평균 필터라 하고 중심 값을 강조하기 위해 필터의 중심 계수를 주변 계수보다 더 크게 하여 사용하며, 고주파 통과 필터는 저주파 차단 필터라고도 하며 신호 성분 중 고주파 성분은 통과시키고 저주파 성분은 차단하는 필터이다. 고주파 통과 필터는 흐려진 영상을 개선하는 첨예화(sharpening) 작업을 수행하며 세밀한 부분을 강조하는 데 쓰인다. 고주파 통과 필터에서 사용되는 마스크는 마스크의 중심에서는 양의 값을 가지며 바깥 경계는 음의 값을 가진다. 또한 고주파 통과 필터 영상은 원 영상에서 저주파 통과 필터링으로 얻은 영상을 빼는 방법으로도 얻을 수 있다. 고주파 필터는 세부 영상을 강조하지만 낮은 주파수에 해당하는 성분의 손실도 발생시킨다. 따라서 고주파 지원 필터링 영상은 원래 영상의 밝기를 증가시킨 후 저주파 영상을 감소시키는 고주파 강조(high-boost) 필터를 사용한다. 고주파 강조 필터는 다음과 같이 나타낼 수 있다.

$$\text{High-Boost} = \text{HB Original} - \text{Lowpass},$$

여기서 HB는 원 영상을 증폭 시키는 값이며 고주파 강조 필터의 가중치이다 HB가 1인 경우 고주파 강조 필터를 통과한 영상은 고주파 통과 필터링의 결과와 같으며 $\text{HB} > 1$ 일 때에는 원래의 영상에 고주파

통과 필터링의 영상이 더해지는 것과 같은 효과를 가진다. 이를 통해 저주파 성분 중 일부를 복원하게 된다. 고주파 강조 필터는 영상에 대한 한 번의 컨볼루션 연산으로 사용할 수 있으며 사용되는 마스크의 중심 계수는 필터 마스크의 중심에 위치한 계수 값을 의미하고 중심 계수 값을 제외한 필터 마스크 내의 모든 계수 값은 $-1/9$ 이다. 따라서 중심 계수 값 $\omega = 9HB - 1$ 이다. 고주파강조 필터는 HB가 증가할 때 원 영상의 배경을 더 많이 반영하므로 필터링된 영상은 원 영상을 더 많이 반영하게 된다. 따라서 낮은 주파수가 많이 존재하는 영상에 고주파 강조 필터를 사용하면 명도의 차이가 커지면서 영상이 좀 더 밝아 보이는 효과를 가져온다.

비선형 공간 필터는 이웃하는 화소들을 토대로 비선형적 연산을 사용하는 필터이다. 비선형 공간 필터에는 마스크 내의 화소 값들의 중위수를 사용하는 중위수 필터가 대표적인 비선형 공간 필터이다. 중위수 필터는 이웃 화소들을 순서대로 정렬 한 후에 중위수를 영상 값으로 선택하는 필터이며, 영상에 급격한 변화가 있는 임펄스 잡음을 제거하는데 사용된다. 중위수 필터는 다음과 같이 나타낼 수 있다.

$$\hat{f}(x, y) = \text{median}_{(s,t) \in S_{xy}} \{g(s, t)\},$$

여기서 g 는 오염된 영상, \hat{f} 는 중위수 필터를 통하여 추정된 원 영상, S_{xy} 는 중심이 (x, y) 인 필터의 영역을 말한다. 중위수 필터는 평균 필터를 이용한 선형 공간 필터 방법에 비해 평활화가 적고 특징들의 경계를 잘 보존하는 장점이 있으나 잡음의 세기가 강하거나 작을 경우에는 영상의 특징들을 보존하지 못하게 되므로 잡음의 정도에 따라 적절히 사용해야 할 것이다.

2.2. 비모수 베이지안

임의의 분포를 가진 자료에 대한 모수적인 접근 방법은 한계가 존재한다. 일반적으로 통계적 추론은 모수적인 가정을 토대로 진행되지만 이러한 가정이 어긋날 경우 추론의 결과에 대해 의문을 제기할 수 밖에 없다. 그러므로 이러한 임의 자료에 대한 모형은 표준 요소(위치모수, 척도모수 등)들이 혼재되어 있다는 혼합모형을 기반으로 한다 (Silverman, 1986). 임의의 분포에 대한 모형은 밀도추정이 주류를 이루고 있으며 이러한 밀도추정을 위한 혼합모형은 커널 방법, 비모수적 최대우도법 (Lindsay, 1983) 또는 디리클레 과정 베이지안 방법 (Ferguson, 1983)들이 대표적이다. 특히 디리클레 과정 베이지안 방법은 모집단의 분포함수에 대하여 특정 형태를 가정하지 않는 비모수적 방법에 가까우며 커널 방법에서 발생할 수 있는 국소적 혹은 대역적 평활화 문제, 평활화 모수의 추정에 대한 문제 등 많은 제한 사항들에 대한 대안이 될 수 있다. 디리클레 과정은 확률과정의 경로 함수들이 확률측도로 표현되는 확률과정이다. 확률과정은 함수 공간에 대한 분포이며 경로 함수들은 그 분포로부터 생성된 임의의 함수가 된다. 디리클레 과정의 경로에는 함수 공간이 구체적으로 확률측도 공간으로 정의되며 이 확률측도 공간에 대한 분포가 디리클레 과정이다. 따라서 디리클레 과정은 확률공간 Θ 에 대한 분포이므로 디리클레 과정에서 생성된 표본들은 임의의 분포를 따른다. 즉, 디리클레 과정은 무한 차원에서 정의된 일반화 디리클레 분포(general Dirichlet distribution)라고 할 수 있다. 디리클레 과정은 기반분포(baseline distribution) G_0 가 임의의 분포 G 의 기대되는 분포라는 가정에 따라 출발하며 중심화 모수 α 를 가지고 있다. α 는 임의의 분포 G 가 기반분포 G_0 를 중심으로 얼마나 분포되어 있는지를 결정하는 모수이며 α 가 커질수록 임의의 분포 G 는 G_0 중심에 분포하게 된다. α 를 정도(precision)모수라고 하며 $\alpha \rightarrow \infty$ 이며 결과적으로 G_0 가 G 가 되며 디리클레 과정은 모수적인 모형이 된다. G_0 받침(support)에서의 분할인 A_1, \dots, A_M 에 대한 확률 벡터 $\{G(A_1), \dots, G(A_M)\}$ 은 $\{\alpha G_0(A_1), \dots, \alpha G_0(A_M)\}$ 를 모수벡터로 갖는 디리클레 분포를 따른다. y_i 가 임의의 모수 θ_i, ϕ_i 를 가진 분포로부터 추출된 표본이라면 $f(y_i|\theta_i, \phi_i)$ 로 나타낼 수 있으며 여기서 θ_i 보다 ϕ_i 에 대한 불확실성이 더 크다고 가정한다 (Escobar와 West, 1998). 즉 θ_i 는 위치모수, ϕ_i 는 척도모수로 생각할 수 있다(θ_i 와 ϕ_i 에 대한 사전분포로 디리클레

x_1	x_6	x_{11}	x_{16}	x_{21}
x_2	x_7	x_{12}	x_{17}	x_{22}
x_3	x_8	x_{13}	x_{18}	x_{23}
x_4	x_9	x_{14}	x_{19}	x_{24}
x_5	x_{10}	x_{15}	x_{20}	x_{25}

Figure 2.1. 5×5 size filter about observed data D .

과정을 사전분포로 가정한다). 기반분포 G_0 는 (θ_i, ϕ_i) 에 대한 후보 값들이 추출되는 분포로 가정하므로 임의의 분포 G 는 디리클레 과정을 사전분포로 하며 이 디리클레 과정의 모수는 α, G_0 이다. G 가 주어졌을 때 모수 (θ_i, ϕ_i) 는 G 에서 부터 독립적으로 추출되며 (θ_i, ϕ_i) 가 주어졌을 때 y_i 는 (θ_i, ϕ_i) 를 모수로 하는 다변량 정규분포를 따른다. 또한 G_0 의 초모수들을 $\gamma = (M, S, B, \eta)$ 라고 하면 다음과 같이 나타낼 수 있다.

$$\begin{aligned} y_i &\sim N(\theta_i, \phi_i) \\ (\theta_i, \phi_i) | G &\sim G \\ G | \alpha, \gamma &\sim \text{DP}(\alpha G_0(\gamma)) \end{aligned}$$

특히 Ishwaran과 James (2002)는 K 개 정규분포가 혼합되어 있다는 정규혼합 모형을 무한 디리클레 과정으로 매우 정확하게 근사시켰다. θ_i 에 대한 가장 적합한 값인 θ_k^* 는 특징이 K 개로 분할되는 확률측도 공간에서 p_k 의 확률로 추출된 값이다. 여기서 추출 확률 p_k 는 정도모수 α 에 의해서 결정된다. p_k 는 혼합의 가중치로 생각할 수 있다. 정도모수 α 는 집락화되는 특징들의 최대 개수로 부여하는 것이 일반적이며 Ohlssen 등 (2007)는 K 번째 가장 확률 p_k 의 기댓값에 근거하여 특징의 갯수 K 의 근삿값을 다음과 같이 제시하였다.

$$K \approx 1 + \log(E(p_K)) / \log \left[\frac{\alpha}{1 + \alpha} \right].$$

이를 토대로 α 를 적절히 선택함으로써 K 의 개수를 결정할 수 있다. K 의 개수는 자료의 본질적인 특성을 반영해야 할 것이다. 동질성을 가정할 수 있는 경우에는 K 의 개수는 자료의 크기보다 작을 것이고 이 경우에는 평활화가 주 목적이라고 할 것이다. 반대로 이질적인 자료들의 경우 자료 자체의 특징들을 모형에 모두 포함시켜야하므로 K 의 개수는 자료의 크기 n 과 유사할 것이다.

2.3. 디리클레 정규 혼합모형

일반적인 정규 혼합모형에서는 임의의 자료 x_i 가 독립적인 정규분포를 따른다고 가정한다. 즉, $X_i | \theta_i \sim N(\mu_i, V_i)$ 이다. 여기서 $\theta_i = (\mu_i, V_i)$ 이며 사전분포는 $R \times R^+$ 를 받침으로 하는 측도 공간에서 정의된다. 관측 자료 $D = \{x_1, \dots, x_n\}$ 은 X_i 의 관측 값 x_i 로 구성되며, 각 관측 값들은 x_i 들은 p 개 존재하므로 D 는 $p \times n$ 의 차원을 가진다. 본 연구에서는 필터 마스크 내의 화소 값들을 시열로 재배열하여 관측 자료 D 로 사용하였으며, 만약 중심의 화소 (x, y) 가 x_{13} 이고 5×5 크기의 필터 마스크를 사용한다면 Figure 2.1과 같이 나타낼 수 있다. 이 경우 D 의 차원은 1×25 이다.

X_{n+1} 은 정규 혼합분포를 따르므로 $X_{n+1} \sim N(\mu_{n+1}, V_{n+1})$ 이고 X_{n+1} 은 $(\theta_{n+1} | D)$ 에 대한 사후예측분포이다. 이 모형은 관측 값 D 의 모수인 θ_i 를 강하게 지지하고 결과적으로 관측 값의 특징이 반영된 사

후분포를 생성하므로 표본 공간상의 국소적인 특징과 전역적인 특징을 모형이 가질 수 있도록 한다. 즉, 관측 값들의 정보를 나타내는 모형을 생성할 수 있게 한다. G 는 $R \times R^+$ 를 받침으로 하는 임의 분포라 하면 θ 는 G 를 사전분포로 갖는다. 만약 G 가 디리클레 과정을 따른다고 한다면 관측 값들은 디리클레 정규 혼합모형(Dirichlet normal mixture)을 따른다고 할 수 있다 (Escobar, 1994; Ferguson, 1983; West, 1990). 임의의 분포는 G 는 디리클레 과정을 따르며 $G \sim DP(\alpha G_0)$ 이다. 여기서 α 는 $\alpha > 0$ 으로 정의되며 G_0 에 대한 측도이면서 디리클레 과정의 초모수이다. 또한 G_0 는 $R \times R^+$ 를 받침으로 가지면서 위치모수와 척도모수에 대한 이변량분포라고 할 수 있다. θ 에 대한 $G_0(\theta)$ 는 $E(G(\theta)) = G_0(\theta)$ 가 성립한다. $\text{Var}(G(\theta)) = G_0(\theta)(1 - G_0(\theta))/(\alpha + 1)$ 이므로 α 는 정도모수이고 G_0 를 기댓값으로 갖는 사전 분포 G 의 퍼짐의 정도를 결정하는 모수이다. 따라서 α 가 커질수록 $G(\theta)$ 의 분산은 작아지게 된다. 만약 G 에서 θ 의 표본을 n 개를 생성할 때 θ_i 는 다음과 같이 나타낼 수 있으며

$$(\theta_i | \theta^{(i)}) \sim \alpha \alpha_{n-1} G_0(\theta_i) + \alpha_{n-1} \sum_{j=1, j \neq i}^n \delta_{\theta_j}(\theta_i).$$

추가 생성되는 θ_{n+1} 에 대한 분포는 다음과 같다.

$$(\theta_{n+1} | \theta^{(n+1)}) \sim \alpha \alpha_n G_0(\theta_{n+1}) + \alpha_n \sum_{i=1}^n \delta_{\theta_i}(\theta_{n+1}),$$

여기서 $\theta^{(i)} = \{\theta_i, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_n\}$, $\theta^{(n+1)} = (\theta_1, \dots, \theta_n)$ 이고 $a_n = 1/(\alpha + n)$ 이다. $\delta_{\theta_i}(\theta)$ 는 $\theta = \theta_i$ 에서의 질량이다. 디리클레 과정에서 α 가 크면 클수록 θ 가 기반 분포에서 추출될 가능성이 높아지고 α 가 작을수록 기반 분포에서 추출될 가능성이 낮아진다. α 는 디리클레 분포를 사용할 때 기반 분포의 지지력을 나타낸다고 할 수 있다. α 는 자료의 크기를 토대로 추측할 수 있으며 임의의 분포 G 로부터 n 개의 α 가 추출 되었다면 θ_{n+1} 는 αa_n 의 확률로 G_0 에서 추출된 값이거나 다른 n 개의 θ_n 들로부터 균등한 확률로 추출된 것이다. 처음 n 개의 값들은 $(\theta_i | \theta^{(i)}) \sim \alpha a_{n-1} G_0(\theta_i) + a_{n-1} \sum_{j=1, j \neq i}^n \delta_{\theta_j}(\theta_i)$ 의 분포를 가지며 이 분포를 토대로 θ 는 $u \leq n$ 개의 특징들로 축소된다 (Antoniak, 1974). θ 로부터 u 개의 특징들로 구성된 새로운 θ 를 $\theta^* = (\theta_1^*, \dots, \theta_u^*)$ 이라고 하면 각 θ^* 들은 각각 n_j 번씩 발생할 것이며 $\sum_{j=1}^u n_j = n$ 이다. 추가 생성된 θ_{n+1} 에 대한 조건부 분포는 다음과 같다.

$$(\theta_{n+1} | \theta) \sim \alpha a_n G_0(\theta_{n+1}) + a_n \sum_{j=1}^u n_j \delta_{\theta_j^*}(\theta_{n+1}),$$

여기서 u 은 $E(u | \alpha, n) \approx \alpha \log(1 + n/\alpha)$ 이므로 α 는 n 에 비하여 상대적으로 작은 값으로 설정해야 하며 표본수가 50에서 250개 수준이라면 α 는 한 자리의 값이 바람직할 것이다. α 에 대한 사전분포가 감마분포이므로 $\alpha \sim \Gamma(a_0, b_0)$ 이며 West (1992)는 $P(\alpha | u, n)$ 의 분포가 다음과 같이 감마분포로 근사함을 보였다.

$$P(\alpha | u, n) \approx \Gamma(a_0 + u - 1, b_0 + \gamma + \log(n)),$$

여기서 γ 는 오일러(Euler) 상수이고 만약 n 이 크다면 $u = o(\log(n))$ 이다.

G_0 를 일반적으로 쓰이는 모형인 정규-역 위사트(normal-inverse Wishart)모형으로 설정하면 다음과 같이 나타낼 수 있다.

$$\begin{aligned} (\mu_i | M, B) &\sim N(M, B), \\ (V_i^{-1} | s, S) &\sim W_p(s, (sS)^{-1}), \\ M &\sim N(M; a, A), \quad S \sim W_p(q; q, q^{-1}R), \quad B^{-1} \sim W_p(B^{-1}; c, (cC)^{-1}), \quad \alpha \sim \Gamma(\alpha; a_0, b_0). \end{aligned}$$

만약 $p = 1$ 이면 $B^{-1} \sim \Gamma(B^{-1}; c, (cC)^{-1})$ 이며, 자료 D 가 주어졌을 때 θ 의 결합 사후분포는 다음과 같다.

$$P(\theta^{(n+1)}|D) = c \prod_{i=1}^n f(x_i|\theta_i) \frac{\alpha G_0(\theta_i) + \sum_{j < i} \delta_{\theta_i}(\theta_j)}{\alpha + i - 1},$$

여기서 c 는 양수인 표준화 상수이고 $f(x_i|\theta_i)$ 는 $N(\mu_i, V_i)$ 에 대한 x_i 의 우도함수이다. 주어진 자료에 대한 사후예측분포는 다음과 같다.

$$P(X_{n+1}|D) = \int P(X_{n+1}|\theta^{(n+1)}) P(\theta^{(n+1)}|D) d\theta^{(n+1)}$$

$P(X_{n+1}|\theta^{(n+1)})$ 는 $\int P(X_{n+1}|\theta_{n+1}) P(\theta_{(n+1)}|\theta^{(n+1)}) d\theta_{n+1}$ 를 계산하면 얻을 수 있으며 다음과 같이 나타낼 수 있다.

$$P(X_{n+1}|\theta^{(n+1)}) = \alpha a_n \int P(X_{n+1}|\theta) dG_0(\theta) + a_n \sum_{j=1}^u n_j f(X_{n+1}|\theta_j^*),$$

$$P(\theta_{n+1}|\theta) = \alpha a_n G_0(\theta_{n+1}) + a_n \sum_{j=1}^u \delta_{\theta_j^*}(\theta_{n+1}),$$

여기서 $f(X_{n+1}|\theta)$ 는 $\theta = (\mu, V)$ 를 모수로 하는 정규밀도함수이다. 마코프 연쇄(Markov chain)를 사용하면 주어진 자료에 대한 θ 의 사후예측분포인 $P(\theta^{(n+1)}|D)$ 에서 θ^{n+1} 를 얻을 수 있으며 표집된 θ 들을 이용하여 몬테카를로 방법으로 사후예측분포 $P(X_{n+1}|D)$ 를 얻을 수 있다. 이를 위해 Neal (2000)과 Müller 등 (1996) 등은 완전 조건부 분포로부터 깃스 표집기를 이용하는 방법을 제안하였다. $\theta_i|\theta^{(i)}$ 에 대한 완전 조건부 분포는 다음과 같다.

$$(\theta_i|D, \theta^{(i)}, M, S, B, \alpha) \sim w_{i0} G^{(i)}(\theta_i) + \sum_{i=1}^{u_j} w_{ij} \delta_j^{i*}(\theta_i),$$

여기서 u_i 는 $\theta^{(i)}$ 에서 구분되는 θ 들의 개수이며 θ_j^{i*} 는 $\theta^{(i)}$ 에서 u_i 개로 축소된 것이다. n_{ij} 는 $\theta^{(i)}$ 에서 같은 θ_j^{i*} 로 분류된 θ 의 개수이다. $G^{(i)}$ 는 우도함수 $f(x_i|\theta_i)$ 와 기반분포 G_0 에 의해 업데이트(update)된 θ_i 의 사후분포이다. 특히 Neal (2000)과 Müller 등 (1996)는 비공액 사전분포를 사용하는 깃스 표집 방법을 제안하였다. 이 방법은 설정 변수 l 를 사용하며, 설정(configuration) 변수는 1에서 n 사이의 정수 값을 가진다. 설정 변수는 관측값 x_i 와 $x_{i'}$ 가 같은 모수 $\theta_i = \theta_{i*} = \theta_j^*$ 를 가진다면 같은 값 $l_i = l_{i'} = j$ 을 가진다. Neal (2000)과 Müller 등 (1996)가 제안한 깃스 표집 방법은 설정 변수에 의해 θ_j^* 가 영향을 받으므로 설정 변수를 요인으로 하는 일원-배치(one-way)의 형태이다. 모든 모수와 초모수에 대한 사후분포는 다음과 같이 나타낼 수 있다.

$$P(\theta^*, M, S, B|D, l, u, \alpha) = \prod_{j=1}^k \left\{ \prod_{i:l_i=j} N(x_i; \mu_j^*, V_j^*) \right\} N(\mu_j^*; M, B) W_p(V_j^{*-1}; s, (sS)^{-1}).$$

또한 μ_j^* 와 V_j^{*-1} 의 조건부 사후분포는 각각 다음과 같다.

$$\begin{aligned} (\mu_j^*|D, V_j^*, M, S, \alpha, l, u) &\sim N(\mu_j^*; M_j, T_j) \\ (V_j^{*-1}|D, \mu_j^*, M, S, B, \alpha, l, u) &\sim W_p(V_j^{*-1}; s+n, \bar{S}_j) \end{aligned}$$

여기서

$$\begin{aligned} T_j^{-1} &= B^{-1} + n_j V_j^{*-1}, \\ M_j &= T_j \left(B^{-1} M + n_j V_j^{*-1} \left(\frac{1}{n_j} \right) \sum_{i:l_i=j} x_i \right), \\ \bar{S}_j^{-1} &= sS + \sum_{i:l_i=j} (x_i - \mu_j^*) (X_i - \mu_j^*)^T \end{aligned}$$

로 각각 나타낼 수 있다. 완전한 l 과 축소된 θ^* 가 주어졌다면 초모수들의 사후분포는 다음과 같이 나타낼 수 있다.

$$\begin{aligned} (M|D, \theta^*, S, B, \alpha, l, u) &\sim N(M; \hat{a}, \hat{A}), \\ (S|D, \theta^*, M, B, \alpha, l, u) &\sim W_p \left\{ S; q + su, \left(qR^{-1} + s \sum_{j=1}^m V_j^{*-1} \right)^{-1} \right\}, \\ (B^{-1}|D, \theta^*, M, S, \alpha, l, u) &\sim W_p \left\{ B^{-1}; c + u, \left(cC + \sum_{j=1}^u (\mu_j^* - M)(\mu_j^* - M)^T \right)^{-1} \right\}, \end{aligned}$$

여기서 \hat{A} 와 \hat{a} 는 각각 다음과 같이 나타낼 수 있다.

$$\hat{A} = A^{-1} + uB^{-1}, \quad \hat{a} = \hat{A} \left(A^{-1}a + uB^{-1} \left(\frac{1}{u} \right) \sum_{j=1}^u \mu_j^* \right)$$

정도모수 α 에 대한 사후분포는 Escobar와 West (1995), West와 Cao (1993)가 다음과 같이 제안하였으며

$$(\alpha|D, \theta^*, M, S, B, l, u, \eta) \sim \pi_1 \Gamma(a_0 + u, b_0 - \log(\eta)) + \pi_2 \Gamma(a_0 + u - 1, b_0 - \log(\eta)),$$

여기서 π_1, π_2 는 다음과 같다.

$$\pi_1 = \frac{a + u - 1}{a_0 + u - 1 + n(b_0 - \log(\eta))}, \quad \pi_2 = 1 - \pi_1$$

는 0과 1사이에서 존재하는 η 의 사후분포는 다음과 같은 베타분포이다.

$$(\eta|D, \theta^*, M, S, B, l, u, \alpha) \sim \beta(\alpha + 1, n).$$

또한 기반분포와 자료에 대한 반영 비율 w_{ij} 는 l 의 사후확률이며 다음과 같이 나타낼 수 있다.

$$w_{ij} = P(l_i = j | D, \theta^*, M, S, B, l^{(i)}, u_i) = \frac{cn_{ij}}{n-1+\alpha} N(x_i; \mu_j^{i*}, V_j^{i*}),$$

여기서 $l^{(i)} = (l_1, \dots, l_{i-1}, l_{i+1}, \dots, l_n)$ 이고 c 는 정규화 상수이다. $j = 0$ 을 가진다는 것은 기반분포에서 θ 를 추출했다는 것이므로 w_{i0} 은 다음과 같이 나타낼 수 있다.

$$w_{i0} = P(l_i = 0 | D, V_j^*, M, S, B, l^{(i)}, u_i) = \frac{c\alpha}{n^*(n-1+\alpha)} N(x_i; M, V_{u+1}^* + B),$$

여기서 V_{u+1}^* 은 G_0 의 위샷트 분포에서 새롭게 생성된 값이고 n^* 는 Neal (2000)이 제안한 보조(auxiliary) 모수이며, n^* 는 보다 빠른 속도로 MCMC를 수행시켜 준다. n^* 는 3 또는 4를 사용하며 (Griffin, 2010) 본 연구에서는 3을 사용하였다. 모든 모수들을 $\phi = (\theta^*, M, S, B, \alpha, l, u)$ 이라 하고 새로운 ϕ_r ($r = (1, \dots, N)$)를 Neal (2000)이 제안한 MCMC 알고리즘8(algorithm 8)을 통하여 얻는 과정은 다음과 같다.

- (1) 새로운 환경 변수를 기존의 ϕ 에 기반한 w_{ij} 와 w_{i0} 를 이용하여 생성한다.
- (2) (1)에서 생성된 새로운 환경 변수를 이용하여 새로운 모수 θ^* 를 $(\mu_j^*|D, V_j^*, M, S, \alpha, l, u)$ 와 $(V_j^{*-1}|D, \mu_j^*, M, S, B, \alpha, l, u)$ 에서 생성한다.
- (3) (2)에서 생성된 새로운 θ^* 를 토대로 새로운 초모수를 $(M|D, \theta^*, S, B, \alpha, l, u)$, $(S|D, \theta^*, M, B, \alpha, l, u)$, $(B^{-1}|D, \theta^*, M, S, \alpha, l, u)$, $(\eta|D, \theta^*, M, S, B, l, u, \alpha)$, $(\alpha|D, \theta^*, M, S, B, l, u, \eta)$ 를 이용하여 생성한다.

이러한 과정을 통해 추출된 새로운 모수 값으로 모든 모수들에 대한 사후분포 $P(\phi|D)$ 를 얻을 수 있고 이 사후분포에서 새롭게 생성된 ϕ_r 을 몬테 카를로 방법을 이용하면 주어진 자료의 사후예측분포를 얻을 수 있다. 특히 Tierney (1994)는 MCMC방법을 사용한 이 사후예측분포가 수렴함을 보였으며 사후예측분포는 다음과 같이 나타낼 수 있다.

$$P(X_{n+1}|D) \simeq N^{-1} \sum_r P(X_{n+1}|\phi_r).$$

Jara 등 (2018)은 프로그램 R에서 위의 알고리즘을 수행하는 ‘DPpackage’를 제안하였으며 구체적으로 ‘DPpackage’내의 ‘DPdensity’ 함수를 사용하면 알고리즘을 수행 할 수 있다.

B^{-1} 를 큰 값으로 설정한다면 V 가 작아지게 하므로 추정되는 결과를 하나의 특징으로 묶일 가능성이 높아지며, B^{-1} 를 작은 값으로 설정하면 자료의 많은 특징들을 부각시켜 다봉형 분포가 생성될 가능성이 높아진다. V 에 대한 사전정보가 충분하다면 B^{-1} 를 작게 설정하고 그렇지 않다면 크게 설정하는 것이 바람직하며, 사전정보가 많지 않다면 초모수 c 와 $(cC)^{-1}$ 를 큰 값으로 설정해야한다. 영상의 잡음 수준에 대한 사전정보가 충분하지 않다고 가정하고 c 와 $(cC)^{-1}$ 에 대한 값으로 100, 1000을 모두 고려해 본 결과 1000이 가장 좋은 결과를 보여 c 와 $(cC)^{-1}$ 를 각각 1000으로 설정하였다. a 와 A 는 필터 마스크내 영상 값들의 표본평균 $\bar{x} = (1/n) \sum_{i=1}^n x_i$ 과 표본분산 $\sigma^2 = (1/(n-1)) \sum_{i=1}^n (x_i - \bar{x})^2$ 으로 설정하였으며 V 와 S 대한 초모수 중 하나인 s 와 q 는 필터 마스크 내의 화소의 수로 하였으며 5×5 크기의 필터 마스크의 경우 $q, s = 25$ 이다. 마지막으로 $q^{-1}R$ 는 필터 내 영상의 표본분산의 역수 $1/\sigma^2$ 로 설정하였다. 정도 모수 α 는 필터 마스크 내의 영상이 가진 특징들을 반영해야 하므로 $a_0 = 1$, b_0 는 필터 마스크 내의 특징 개수의 역수로 설정하였다. 영상이 특징 없는 평활한 영역이면 화소의 밝기값들은 유사한 밝기 값으로 나타날 것이고 잡음에 의해 영향을 받더라도 가장 큰 자리수의 밝기 값은 변화가 상대적으로 둔감할 것이다. 밝기 값들을 특정 자리수에서 반올림하면 단 하나로 묶여지게 될 것이다. 그러므로 $1/b_0$ 는 필터 내의 밝기 값들을 특정 자리수로 반올림하였을 때 구분되는 밝기 값들의 개수로 설정하였다. 특징의 개수를 많이 파악하기 위해서는 낮은 자리수에서 반올림을 하도록 설정하고 반대인 경우에는 높은 자리수에서 반올림을 하도록 설정하였다.

Griffin (2010)은 초모수 α 보다 B^{-1} 가 미지의 분포에 대한 추정에 영향을 미치는것을 보였으며 α 가 크더라도($\alpha = 50$) $B^{-1}(B^{-1} \leq 20)$ 가 크면 미지의 분포에 대한 추정이 비모수적으로 이루지는 것을 보였다. 본 연구에서는 B^{-1} 의 초모수 c 와 $(cC)^{-1}$ 를 각각 1000으로 설정하였으므로 B^{-1} 는 충분히 크다고 할 수 있으며($E(B^{-1}) = 1000000$), Neal (2000)이 제안한 방법은 비공액 사전분포를 이용하여 MCMC를 구성하므로 미지의 모수에 대한 추정이 모수적이기 보다는 비모수적이고 기반분포의 초모수에 영상의 정보를 나타내는 통계량들을 사용하여 사전정보를 추론에 사용하고자 하였다. 또한 본 연구의 알고리즘에서 사용하는 값은 원 영상에 대한 사후예측 값이며, 새로 표집된 사후예측 값의 분포가 원 영상의 분포와 유사해야 한다. 이를 확인하기 위해 크기 필터 마스크의 영상을 토대로 영상의 표본통계량을 초모수로 사용한 경우(a 를 영상의 표본평균, A 는 영상의 표본분산)와 무정보적(noninformative) 초모수($a = 0, A = 100000$)를 사용한 경우의 사후예측 값을 비교하여 보았다. Figure 2.2는 왼쪽 상단

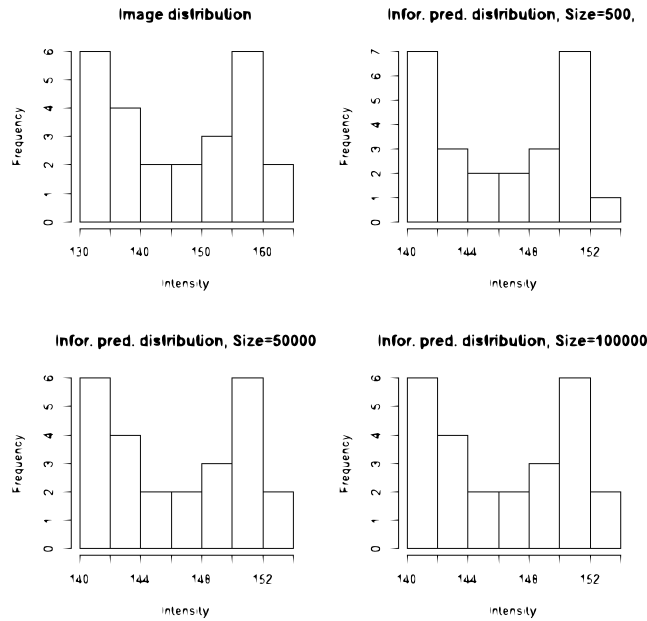


Figure 2.2. Histogram of original values and predictive values using informative hyperparameters.

부터 시계 방향으로, 원 영상 밝기 값의 분포와 연쇄의 크기에 따라 표본 통계량을 사용한 사후예측 값의 분포가 어떻게 변하는지 보여준다. 사후예측 값의 분포가 원 영상의 분포와 유사한 형태를 보여야 적절하므로 표본 통계량을 사용한 경우에는 연쇄의 크기에 관계없이 원 영상의 분포를 잘 반영하고 있다. Figure 2.3은 무정보적 초모수를 사용한 사후예측 값의 분포를 나타낸다. 무정보적으로 초모수를 설정할 경우 연쇄의 크기가 충분히 크더라도 원 영상의 분포와 다소 차이가 존재하므로 영상의 정보를 반영하도록 초모수를 설정하는 것이 바람직할 것이다.

3. 영상 특징의 베이지안 추론을 적용한 적응적 필터

3.1. 적응적 필터

적응적 필터는 영상에 포함된 잡음 수준을 지역적으로 처리하는 필터링 방법이다. 필터 내에 속하는 영상의 밝기 수준을 영상 값들의 평균으로 이해하고 영상에 포함된 잡음 수준을 영상 값들의 분산으로 접근하여 영상 전체의 분산 수준과 필터 내의 분산 수준을 비교한다. 따라서 적응적 필터는 다음과 같이 나타낼 수 있다.

$$X_{g(x,y)} = X_{h(x,y)} - \frac{\sigma_t^2}{\sigma_h^2} (X_{h(x,y)} - \bar{X}_h),$$

여기서 σ_t^2 은 영상 전체의 잡음 분산이며 σ_h^2 는 필터 내의 분산, \bar{X}_h 는 필터 내의 평균이다. $\sigma_t^2 = 0$ 이면 필터는 원 영상을 반환하며, 필터 내의 분산 수준보다 전체의 분산 수준이 더 높다면 필터는 필터 내의 평균값을 반환한다. 일반적으로 $\sigma_t^2 \leq \sigma_h^2$ 이나 $\sigma_t^2 > \sigma_h^2$ 가 발생할 경우에 $\sigma_t^2/\sigma_h^2 = 1$ 로 하여 필터 내의 평균값을 출력하고 밝기 수준이 음의 값을 갖지 않도록 하였다. 따라서 적응 필터는 원 영상의 특징들을 잘 살려내고 특히, 가우시안 잡음 제거에 뛰어나지만, 임펄스 잡음 제거에는 적합하지 않으며 σ_t^2 에 대해 정확히 알고 있을 때만 사용이 가능한 단점이 있다.

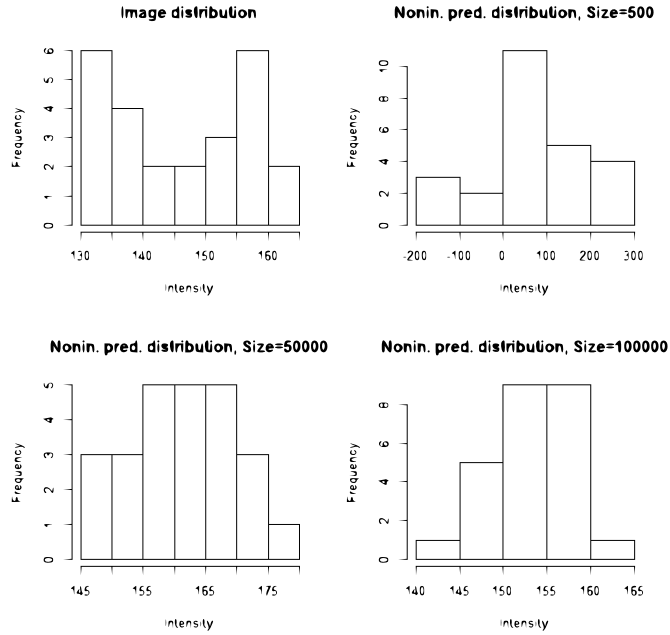


Figure 2.3. Histogram of original values and predictive values using noninformative hyperparameters.

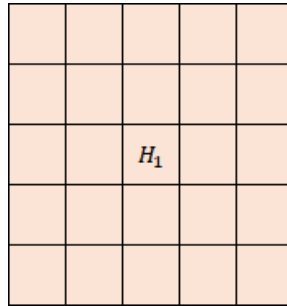


Figure 3.1. 5×5 H_1 filter.

3.2. 비모수 베이지안 적응적 필터

영상에 포함된 특징들은 영상의 좌푯값에 따라 구분되어 나타나므로 영상의 특징은 지역적으로 비슷한 값들이 나타날 것이며 이러한 지역적 특색을 고려하는 것이 중요하다. 기존의 필터는 지역적 특색을 고려하지만, 필터의 크기에 비례하여 위치하는 정보만 사용하므로 영상의 특징을 구별하기에는 제한적이다. 따라서 공간 필터에 사용하는 일반적인 필터 이외에도 지역적인 특징을 확인하기 위한 지역적 필터를 고려하자. Figure 3.1과 같이 일반적인 5×5 필터 H_1 를 기준으로 H_1 을 포함하고 H_1 보다 더 큰 필터 H_2 를 고려하자. 예를 들어 H_2 필터가 H_1 필터의 넓이 보다 5배가 더 크다면 H_2 필터의 크기는 25×25 이며, Figure 3.2와 같다. 이를 일반화하여 H_1 필터의 크기를 $w_1 \times w_1$, H_2 필터의 크기를 $w_2 \times w_2$ 라고 하자. 만약 영상에 잡음이 포함된다면 기존의 적응적 필터는 H_1 의 밝기 분산 σ_h^2 와 영상 전체의 밝기 분산 σ_i^2 만을 이용하여 영상에 포함된 잡음을 감소시켰지만 이는 앞서 언급했듯이 영상 전체의 잡음에 대한 정확한 정보를 얻기 어렵다. 일반적으로는 영상 전체의 잡음을 표본분산으로 사용하지만 만약 영

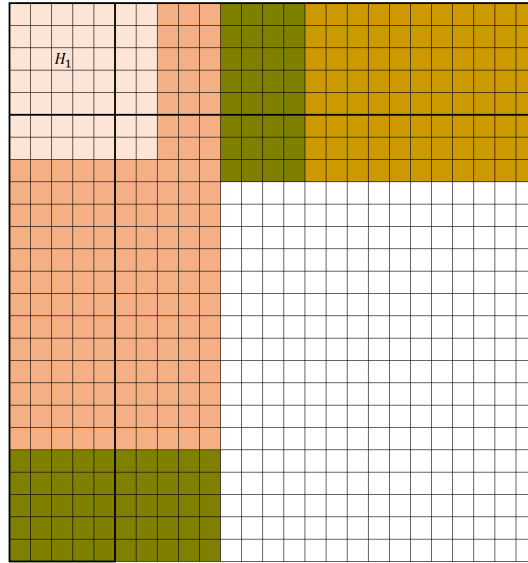


Figure 3.2. 25×25 H_2 filter.

상이 많은 특징을 가지고 있다면 영상의 표본분산은 잡음이 없음에도 불구하고 높은 값을 가질 것이며 $\sigma_h^2 \geq \sigma_t^2$ 이므로 기존의 적응적 필터는 특징들이 열화되어 나타날 것이다. 단순히 전체의 표본분산을 사용하기보다 H_2 의 표본분산을 사용한다면 유사한 특징 내의 분산 수준을 고려하게 되므로 좀 더 영상의 특징을 고려하게 될 것이다. 또한 H_2 의 표본분산 보다는 H_1 의 근방에서의 잡음을 사용한다면 영상의 특징을 잘 표현하게 될 것이다.

H_1 을 기준으로 종 방향의 H_{1_h} 필터와 횡 방향의 H_{1_v} 필터를 고려하자. 위의 예에서 H_1 의 특징은 H_{1_h} 보다 H_{1_v} 에 가깝다고 할 수 있다. 따라서 H_{1_v} 의 표본분산을 사용하는 것이 좀 더 바람직 할 것이다. 디리클레 정규 혼합모형으로 추정된 H_{1_v} 의 사후예측분포를 $P(X_{v+1}|X_v)$ 라고 하고 이 사후예측분포에서 관측값에 대한 조건부 예측 값(conditional predictive ordinate; CPO_i)을 구하면 CPO_i 는 다음과 같이 나타낼 수 있다.

$$CPO_i = P(x_i | x^{(i)}),$$

여기서 $x^{(i)} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$ 이고 P 는 사후예측분포이다. 주어진 영상 필터 내의 사후예측분포로 이를 다시 표현 하면 다음과 같다.

$$CPO_i = P(X_{v+1_i} | X_{v+1}^{(i)}),$$

여기서 $X_{v+1}^{(i)} = (X_{v+1_1}, \dots, X_{v+1_{i-1}}, X_{v+1_{i+1}}, \dots, X_{v+1_n})$ 이다. CPO_i 에 대한 추정량은 MCMC를 통해서 다음과 같이 얻을 수 있다.

$$\widehat{CPO}_i = \left[\frac{1}{T} \sum_{t=1}^T P(X_{v+1_i} | X_{v+1}^{(i)}) \right]^{-1},$$

여기서 $t = (1, \dots, T)$ 이고 MCMC를 통해 추출한 표본의 순번이다. CPO_i 는 다른 모든 자료 $X_{v+1}^{(i)}$ 를 이미 관측하고 난 뒤에 X_{v+1_i} 에 대한 확률을 나타내며 CPO_i 가 작을수록 X_{v+1_i} 가 이상치(outlier)이거

나 다른 관측치에 영향을 미치는 관측치로 생각할 수 있다. 로그-유사 주변 우도(log-pseudo marginal likelihood; LPML)는 CPO_i 를 토대로 산출되며 로그 CPO_i 의 평균이며, LPML은 다음과 같이 나타낼 수 있다.

$$LPML = \frac{1}{n} \sum_{i=1}^n \log(CPO_i).$$

두 모형의 LPML에 대한 비(ratio)는 베이지 요인으로 간주할 수 있으며 유사 베이지 요인(pseudo Bayes Facotor; psBF)은 $LPML_h$ 와 $LMPL_v$ 를 각각 종 방향 필터 마스크와 횡 방향 필터 마스크에서 계산한 LMPL이라 하면 다음과 같다 (Held와 Bovè, 2014).

$$psB_{vh} = \frac{LPML_h}{LPML_v}$$

엄밀한 의미의 베이지 요인은 동일한 자료 X 가 주어졌을 때 계산되나, 본 알고리즘에서는 지역적 필터 내부 영상 값이 주어졌을 때 그것의 사후확률을 비교하여 적절한 영상 값을 선택하고자 하므로 그 의미는 유사하다고 할 수 있다. 영상의 잡음 수준이 낮다면 σ_t^2 는 σ_h^2 에 비해 현저히 작은 값을 가져야 하고 영상의 잡음 수준이 높다면 $\sigma_t^2/\sigma_h^2 = 1$ 이 되어야 할 것이다. Figure 3.2에서 잡음 수준이 낮으면 영상의 특징이 잘 나타나므로 종 필터에서 추정된 분산이 횡 필터에서 나타난 분산보다 더 높을 것이다. 잡음 수준이 높으면 종 필터에서 추정된 분산이 더 클 것이다. 디리클레 혼합 모형은 자료에 포함된 특징들을 묶어주므로 특징들이 적게 포함될수록 사후예측확률은 높게 계산된다. 그러므로 Figure 3.2에서 잡음이 낮은 수준으로 추가된다면 종 필터의 사후확률은 특징들이 더 많으므로 횡 필터보다 작게 된다. 반대로 잡음이 높은 수준으로 추가된다면 분산 수준을 높게 해야 하므로 b_{vh} 가 양수 값으로 계산되는 필터인 종 필터의 분산이 σ_t^2 가 된다. 따라서 제안하는 적응적 필터의 알고리즘은 다음과 같다.

- Step 1. $H_{1(x,y)}$ 와 $H_{2(x,y)}$, $H_{v(x,y)}$ 와 $H_{h(x,y)}$ 를 설정한다.
- Step 2. $H_{v(x,y)}$ 와 $H_{h(x,y)}$ 에서 각각의 사후예측분포 $P(X_{v+1}|X_v)$ 와 $P(X_{h+1}|X_h)$ 을 계산한다.
- Step 3. 베이지 요인 psB_{vh} 를 통하여 영상의 특징들이 구별되기 어려울 정도로 잡음 수준이 높다면 psB_{vh} 가 양수 값으로 계산되는 필터를 선택하고 그렇지 않다면 psB_{vh} 가 음수 값으로 계산되는 종 또는 횡 필터를 선택한다.
- Step 4. $H_{1(x,y)}$ 에서 디리클레 정규혼합 모형으로 사후예측값 $Y_{1(x,y)}$ 을 생성한다. 여기서 생성된 사후예측값은 각각의 사후예측값들의 평균을 사용한다.
- Step 5. σ_t^{*2}/σ_h^2 를 계산한다. σ_t^{*2} 는 $H_{v(x,y)}$ 나 $H_{h(x,y)}$ 의 사후예측값 $Y_{v(x,y)}, Y_{h(x,y)}$ 의 표본분산이다. 즉,

$$\sigma_t^{*2} = \frac{1}{w_1 w_2 - 1} \sum_{x,y}^{w_1, w_2} (Y_{(x,y)} - \bar{Y})^2$$

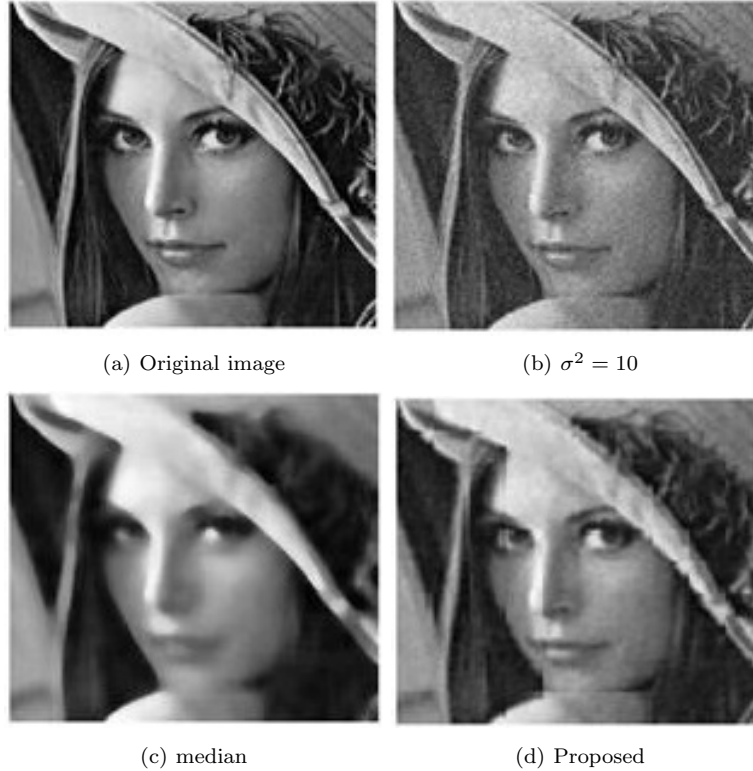
또한, $\sigma_t^{*2}/\sigma_h^2 > 1$ 인 경우에는 $\sigma_t^{*2}/\sigma_h^2 = 1$ 이다. σ_h^2 는 H_1 필터에서 생성된 사후예측값들로 계산된 표본분산이다.

- Step 6. 계산된 σ_t^{*2}/σ_h^2 를 적응적 필터에 적용한다 $Y_{1(x,y)} - \sigma_t^{*2}/\sigma_h^2 (Y_{1(x,y)} - \bar{Y}_1)$.

본 연구에서는 기반 분포에 대한 초모수들을 앞서서 언급한 바와 같이 설정하였다. $(c, (cC)^{-1}) = 1000$, $a = \bar{x}$, $A = \bar{\sigma}^2$, $(q, s) = 25$, $q^{-1}R = 1/A$, $a_0 = 1$, 그리고 b_0 는 필터 마스크 내의 화소값을 특정 자리수에서 반올림하여 구분되는 화소값들의 개수로 하였다. 반올림 자리수는 1의 자리와 100의 자리를 고려하였다. 영상에 잡음이 많이 포함된 경우에는 100의 자리에서 밝기 값을 반올림 하였으며 영상에 잡음이 적게 포함된 경우에는 1의 자리에서 밝기 값을 반올림 하였다.

Table 3.1. Simulation results (peak signal-to-noise ratio)

σ	Proposed filter		Median filter		Adaptive filter	
	Mean	Std	Mean	Std	Mean	Std
10	70.5711	0.0303	69.1863	0.1273	69.9540	0.0245
20	70.6165	0.0438	69.1656	0.3800	69.7316	0.0346
30	69.7420	0.0500	68.4398	0.3252	69.3586	0.0443

**Figure 3.3.** 200×200 Lena images.

3.3. 모의 실험

제안한 알고리즘의 성능을 확인하기 위하여 원 영상에 임의의 잡음이 추가된 잡음 모형을 고려하였다. 잡음이 추가된 오염 모형은 $d(x, y) = f(x, y) + n(x, y)$ 이며 여기서 f 는 원 영상 n 은 잡음을 가르킨다. 잡음은 영상처리 분야에서 일반적으로 많이 사용하는 가우시안 잡음과 균일 잡음이 혼합된 잡음을 사용하였다. 따라서 $n(x, y) \sim N(0, \sigma_n^2) + U(-3\sigma_n, 3\sigma_n)$ 이고 잡음 수준은 $\sigma_n = \{10, 20, 30\}$ 로 하여 각각의 잡음 수준에 따른 알고리즘의 성능을 확인하였다. 또한 지역적 필터는 각각 $w_1 = 5 \times 5$, $w_2 = 25 \times 25$ 로 설정하였다. 사용한 원 영상은 영상처리 분야에서 널리 사용되고 있는 Lena 영상으로 알고리즘의 성능을 확인하였다. 영상처리는 R version 3.4.2를 통해 진행하였으며 DPpackage (Jara 등, 2018)를 사용하였다. MCMC 설정 값은 burn-in은 1,000개, 연쇄의 부분표집은 5번째까지의 표집은 사용하지 않고 6번째의 연쇄에서 표집되는 표본들을 사용하였다. 또한 사용한 표본의 수는 500개로 하였으므로 표집되는 총 표본은 7,000개였다. 모의실험은 잡음수준별로 총 10회씩 반복하였으며 알고리즘의 성능은

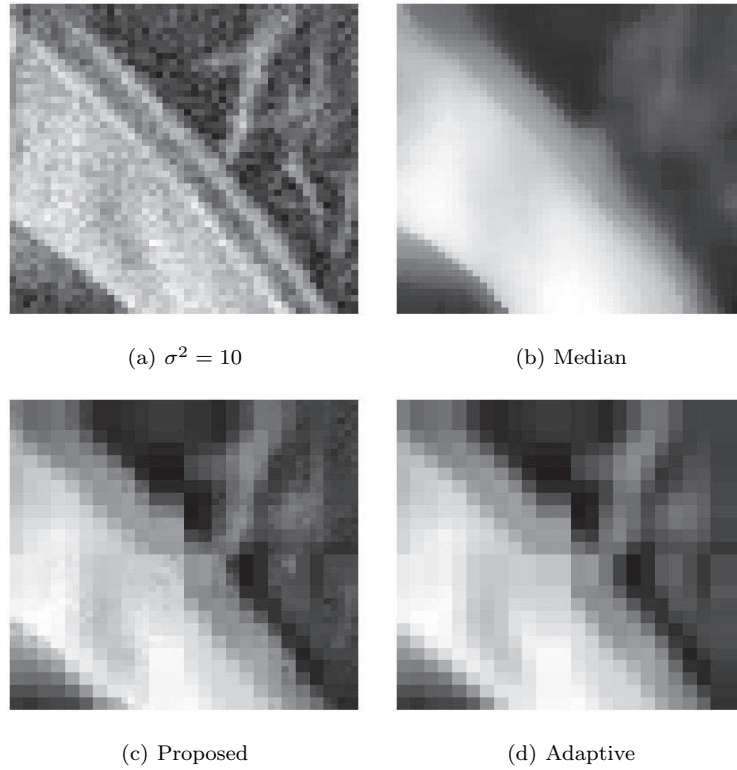


Figure 3.4. 50×50 Lena images.

peak signal-to-noise ratio (PSNR)로 평가하였으며 다음과 같이 나타낼 수 있다.

$$\text{PSNR} = 10 \times \log_{10} \left\{ \frac{\text{MAX}_I^2}{\text{MSE}} \right\},$$

여기서 $\text{MAX}_I = 255$ 이고 mean squared error (MSE)는 O 를 원 영상 \hat{H} 를 각 필터로 처리된 영상이라고 하면 다음과 같이 나타낼 수 있다.

$$\text{MSE} = \frac{1}{N^2} \sum (O - \hat{H})^2,$$

여기서 O 는 원 영상, H 는 각 필터를 통해 처리된 잡음 제거 영상이다. N 은 $N \times N$ 크기의 영상에서의 N 이다. MSE는 오차 제곱합이므로 PSNR 값이 클수록 처리된 영상이 원 영상에 가깝다. 모의실험을 잡음수준별로 10회씩 반복하여 중위수 필터와 기존의 적응적 필터 그리고 제안한 필터의 PSNR 값의 평균과 표준편차를 확인하였다. Table 3.1에서 확인할 수 있듯이 제안한 필터의 PSNR 값의 평균이 다른 필터들에 비하여 모든 잡음 수준에서 더 낮은 값을 보이고 있음을 알 수 있다. 또한 분산이 영상에 작게 포함될수록 제안된 필터의 PSNR 값이 가장 컸다. 한편, 영상에 분산이 큰 잡음이 포함되더라도 중위수 필터나 기존의 적응적 필터보다 PSNR 값이 크게 나타났다. Figure 3.3과 같이 실제로 영상들을 시각적으로 확인해 보더라도 열화가 심하게 발생하는 중위수 필터보다 특징들을 잘 보존하는 것을 알 수 있다. 기존의 적응적 필터와 제안한 필터의 차이를 보다 자세하게 확인하기 위하여 영상의 특정 부분을

확대하여 비교하였다. 확대한 Figure 3.4 영상에서는 기존의 적응적 필터보다 국소적인 특징들을 잘 나타내고 있음을 알 수 있다.

4. 결론

영상에 포함된 잡음을 제거하는 방법들은 잡음의 분포에 대한 가정을 토대로 진행되지만 가정이 위배되거나 충족시킬수 없는 경우 이러한 방법들은 잡음을 원활하게 제거하지 못한다. 영상의 잡음을 제거하는 알고리즘은 영상의 지역적인 특징들을 반영해야하며 지역적 특징들을 열화시키는 경우에는 알고리즘의 성능이 떨어진다. 공간 필터에서 사용하는 필터 마스크는 영상의 지역적 특징을 반영하는 도구이지만 필터 마스크가 연산하는 영역은 소표본에 해당한다. 본 연구에서는 잡음의 분포에 대한 가정을 할 수 없는 경우와 표본의 크기가 작은 경우에도 사용할 수 있는 방법을 고려하였다. 디리클레 정규 혼합모형은 디리클레 과정의 초모수에 따라 임의의 분포를 세밀하게 추정하므로 영상이 가진 특징들을 잘 보존한다. 또한 영상이 가진 특징들을 중심으로 화소 값들을 묶어주는 역할도 하므로 큰 잡음이 포함되었을 때에는 영상의 특징을 살리면서 잡음을 감소시켜준다. 적응적 필터는 영상의 전체의 표본 분산을 기준으로 필터 마스크내의 표본 분산을 비교하는 필터이다. 필터 마스크내의 표본 분산은 전체의 표본 분산보다는 필터 마스크 근방의 표본 분산과 비교하는것이 지역적 특징을 반영해 준다. 디리클레 정규 혼합모형을 토대로 지역적 특징을 반영하는 필터를 사용하여 적응적 필터에 적용하였다. 모의 실험의 결과에서 확인할 수 있듯이 PSNR 값이 기존의 다른 방법들에 비하여 높은 값을 보이고 있으며 Lena 영상의 처리 결과에서도 열화가 적고 잡음이 적절히 제거되고 있음을 알 수 있다. 특히 잡음의 분산이 작을 때 뿐만 아니라 잡음의 분산이 클 때도 기존의 방법에 비해 더 나은 결과를 보이고 있다.

References

- Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems, *The Annals of Statistics*, **2**, 1152–1174.
- Baek, S., Jeong, S., Choi, J. S., and Lee, S. (2015). Effective noise reduction using STFT-based content analysis, *Journal of The Institute of Electronics and Information Engineers*, **52**, 145–155.
- Escobar, M. D. (1994). Estimating normal means with a Dirichlet process prior, *Journal of the American Statistical Association*, **89**, 268–277.
- Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures, *Journal of the American Statistical Association*, **90**, 577–588.
- Escobar, M. D. and West, M. (1998). Computing nonparametric hierarchical models. In *Practical Nonparametric and Semiparametric Bayesian Statistics* (pp. 1–22). Springer, New York.
- Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems, *The Annals of Statistics*, **1**, 209–230.
- Ferguson, T. S. (1983). Bayesian density estimation by mixtures of normal distributions. In *Recent Advances in Statistics*, Academic Press.
- González, R. C. and Woods, R. E. (2008). *Digital Image Processing* (3rd ed), Prentice Hall, NJ.
- Griffin, J. E. (2010). Default priors for density estimation with mixture models, *Bayesian Analysis*, **5**, 45–64.
- Held, L. and Bovè, D. S. (2014). Bayesian Inference. In *Applied Statistical Inference*, Springer, Berlin, Heidelberg.
- Ishwaran, H. and James, L. F. (2002). Approximate Dirichlet process computing in finite normal mixtures: smoothing and prior information, *Journal of Computational and Graphical Statistics*, **11**, 508–532.
- Jara, A., Hanson, T., Quintana, F., Mueller, P., Rosner, G., and Jara, M. A. (2018). Package ‘DPpackage’.
- Kim, Y. H. and Nam, J. (2011). Estimation of the noise variance in image and noise reduction, *The Korean Journal of Applied Statistics*, **24**, 905–914.

- Kim, Y. H. (2012). Adaptive noise reduction algorithm for image based on block approach, *Communications for Statistical Applications and Methods*, **19**, 225–235.
- Lindsay, B. G. (1983). The geometry of mixture likelihoods: a general theory, *The Annals of Statistics*, **11**, 86–94.
- Müller, P., Erkanli, A., and West, M. (1996). Bayesian curve fitting using multivariate normal mixtures, *Biometrika*, **83**, 67–79.
- Müller, P. and Quintana, F. A. (2004). Nonparametric Bayesian data analysis, *Statistical Science*, **19**, 95–110.
- Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models, *Journal of Computational and Graphical Statistics*, **9**, 249–265.
- Ohlssen, D. I., Sharples, L. D., and Spiegelhalter, D. J. (2007). Flexible random-effects models using Bayesian semi-parametric models: applications to institutional comparisons, *Statistics in Medicine*, **26**, 2088–2112.
- Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.
- Song, M., Park, C., and Lee, J. (2003). *Non-Parametric Statistics using S-LINK*, Free Academy, Seoul.
- Tierney, L. (1994). Markov chains for exploring posterior distributions, *The Annals of Statistics*, **22**, 1701–1728.
- West, M. (1990). Bayesian kernel density estimation, Institute of Statistics and Decision Sciences, Duke University.
- West, M. (1992). Hyperparameter estimation in Dirichlet process mixture models. ISDS Discussion Paper# 92-A03: Duke University.
- West, M. and Cao, G. (1993). Assessing mechanisms of neural synaptic activity. In *Case Studies in Bayesian Statistics*. Springer, New York.
- Zhou, M., Chen, H., Paisley, J., *et al.* (2012). Nonparametric Bayesian dictionary learning for analysis of noisy and incomplete images, *IEEE Transactions on Image Processing*, **21**, 130–144.

비모수 베이지안 방법을 이용한 영상 잡음 제거 알고리즘

우호영^a · 김영화^{a,1}

^a중앙대학교 응용통계학과

(2018년 4월 9일 접수, 2018년 6월 14일 수정, 2018년 8월 1일 채택)

요약

영상처리 분야의 중요한 주제인 영상의 잡음 제거 과정은 원래의 순수한 영상이 다양한 원인으로 발생한 잡음에 의해 오염되었을 때 이 잡음을 제거하거나 줄이는 것을 의미한다. 잡음 제거 과정에서는 영상에 추가된 잡음과 원 영상이 가진 고유한 특징들을 구별해내는 것이 중요하며 이에 대한 많은 연구가 진행되고 있다. 적응적 필터와 시그마 필터는 잡음 제거를 위하여 사용하는 대표적인 잡음 제거 필터이며 이 필터들의 효용성은 정확한 잡음 추정에 영향을 받는다. 따라서 본 연구에서는 디리클레 정규 혼합모형을 토대로 영상을 오염시키고 있는 잡음의 분포를 생성하고 이를 토대로 영상의 특징과 잡음을 구별하기 위한 베이지안 방법을 제시한다. 특히 잡음의 분포와 특징의 분포를 구별하기 위해 베이지안 추론을 전개하고 영상에 포함된 잡음을 제거하는 알고리즘을 제시하고자 한다.

주요어: 적응적 필터, 베이지안, 디리클레 정규혼합 모형, 영상처리, 잡음 제거

이 논문은 2017년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No.2017R1D1A1B03031725).

¹교신저자: (06974) 서울특별시 동작구 흑석로 84, 중앙대학교 응용통계학과. E-mail: gogators@cau.ac.kr