

파워마스크를 이용한 영상 핑거프린트 정합 성능 개선

서진수[†]

Improving Image Fingerprint Matching Accuracy Based on a Power Mask

Jin Soo Seo[†]

ABSTRACT

For a reliable fingerprinting system, improving fingerprint matching accuracy is crucial. In this paper, we try to improve a binary image fingerprint matching performance by utilizing auxiliary information, power mask, which is obtained while constructing fingerprint DB. The power mask is an expected robustness of each fingerprint bit. A caveat of the power mask is the increased storage cost of the fingerprint DB. This paper mitigates the problem by reducing the size of the power mask utilizing spatial correlation of an image. Experiments on a publicly-available image dataset confirmed that the power mask is effective in improving fingerprint matching accuracy.

Key words: Image Fingerprinting, Image Hashing, Power Mask, Weighted Hamming Distance, Image Search

1. 서론

방대한 양의 디지털 영상을 대상으로 영상 추천, 영상 데이터베이스 관리, 저작권 보호 등을 빠르고 신뢰성 있게 제공해 줄 수 있는 검색 서비스의 필요성이 커지고 있으며, 콘텐츠 식별 기술인 핑거프린팅은 검색 서비스를 위한 핵심 기술이다 [1-3]. 콘텐츠 식별은 생체 식별에서 사람의 지문, 홍채 등을 이용하여 그 사람을 인식하는 것처럼 콘텐츠의 특징을 이용하여 해당 콘텐츠를 식별하는 기술을 말한다. 이때 사용되는 특징을 핑거프린트(fingerprint) 또는 해시(hash)라고 부른다. Fig. 1과 같이 인식하고자 하는 영상들에서 핑거프린트를 추출하여 핑거프린트 데이터베이스를 만들고 영상의 메타 정보와 연동시키게 된다. 인식하고자 하는 미지의 영상에서 추출된 핑거프린트로 미리 만들어둔 핑거프린트 데이터베

이스를 검색하고 최종 검증 과정을 거쳐 인식하게 된다. 핑거프린팅을 이용한 영상 식별은 P2P/UCC 등을 통한 불법 파일 공유를 막는 필터링, 방송 모니터링, 대용량 영상 라이브러리를 자동으로 태깅(tagging) 또는 인덱싱(indexing) 하는 등 다양하게 응용될 수 있어서 많은 관심을 받아 왔다.

기존의 암호 기반 데이터 해싱 기법의 경우 데이터가 조금만 변화해도 해시 값이 크게 변화하므로, 영상 신호와 같이 압축, 잡음처리, 아날로그-디지털/디지털-아날로그 변환 등의 다양한 신호처리 과정에 대한 강인성이 요구되는 경우에는 적합하지 않다. 콘텐츠 식별에 사용되는 핑거프린트는 간결하면서도 다양한 변환에 대한 강인성을 가지고, 서로 다른 콘텐츠에 대해 차별성을 줄 수 있어야 한다 [1]. 따라서 영상 핑거프린트 추출 및 정합은 간결성, 강인성, 차별성을 두루 고려하여 설계하여야 한다. 영상 핑거프

* Corresponding Author : Jin Soo Seo, Address: (25457) 7 Jukhun-gil, Gangneung, Gangwon-do, Korea, TEL : +82-33-640-2428, FAX : +82-33-646-0740, E-mail : jssseo@gwnu.ac.kr

Receipt date : Sep. 24, 2019, Revision date : Nov. 14, 2019
Approval date : Dec. 16, 2019

[†] Dept. of Electrical Eng., Gangneung-Wonju National University

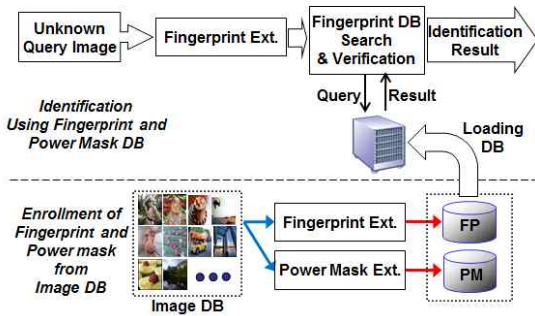


Fig. 1. Overview of fingerprint and power mask extraction for image identification.

린트의 형태는 실수와 이진수 값의 형태를 가질 수 있으며, 본 논문에서는 이진수 형태의 핑거프린트만을 고려한다. 일반적으로 핑거프린트 추출 과정은 영상에서 특징을 추출하고 이진화하는 과정으로 이루어진다. 핑거프린트를 추출하기 위해서 영상 무작위 블락들의 평균[4], 영상의 라돈 변환[1], 영상 블락의 최소 코드[5], 영상 특징점들의 상대적 위치[2], 영상 텐서 변환[6] 등의 특징들이 사용되었다. 특징 이진화는 특징을 직접 양자화하거나 특징의 부호를 취하는 등의 방법이 있다. 각각의 영상에서 얻어진 이진 핑거프린트는 Fig. 1에 주어진 바와 같이 데이터베이스에 저장되며, 검색 대상 미지의 영상으로부터 얻은 핑거프린트와 데이터베이스에 저장된 핑거프린트 간의 정합은 일반적으로 해밍 거리를 이용한다.

영상 핑거프린팅을 통해서 영상을 식별하기 위해서는 원본 영상의 핑거프린트와 다양한 왜곡이 가해진 영상으로부터 얻은 핑거프린트 간의 차이가 작아야 하며, 서로 다른 영상들에서 얻은 핑거프린트들 간의 거리는 충분히 커야한다. 기존 연구들은 영상 식별 성능을 제고하기 위해서 주로 핑거프린트 추출 방법 개선에 관하여 다루었다. 하지만 핑거프린트를 얻는 과정인 특징 추출과 이진화 등에서 필연적으로 정보의 손실이 발생하므로, 이진수 형태의 핑거프린트만으로 영상 식별 성능을 개선하는 것은 한계가 있다. 이를 반영하여 최근 핑거프린트 추출이 아닌 정합 방법을 개선하여 식별 성능을 제고하는 방법들에 대한 연구들이 시도되었으며, 파워 마스크 방법 [7], 비대칭 정합 방법 [8, 9] 등이 제안되었다. 본 논문은 오디오 핑거프린팅에 적용되었던 파워 마스크를 영상 핑거프린트 정합 성능 제고에 활용하는 방안에 대하여 다룬다. 파워 마스크는 원본 콘텐츠로부터

핑거프린트를 추출하여 데이터베이스화하는 과정에서 얻은 정보를 활용하여 각 핑거프린트 비트의 예측 강인도를 추정한 것이다. 일반적으로 파워 마스크 저장에 소요되는 저장 공간의 크기를 줄이기 위해서 추정한 예측 강인도를 이진화하여 강인한 비트와 연약한 비트로 나누어 저장한다. Fig. 1의 영상 식별기에서 검색하고자 하는 미지의 입력 영상에서 핑거프린트를 추출하고 데이터베이스에 있는 핑거프린트와 정합 시에 각 핑거프린트 비트의 파워 마스크 값이 가중치로 사용된다. 파워 마스크는 정합 성능을 향상시키지만 데이터베이스 저장 공간도 늘어나는 문제점이 있다. 기존 논문에서는 이를 해결하기 위해서 파워 마스크를 크기 순서에 따라 이진화하였다 [7]. 하지만 파워 마스크를 이진화하더라도 파워 마스크를 저장하기 위해서 필요한 저장 공간이 핑거프린트 저장 공간의 크기와 같게 된다. 본 논문에서는 영상의 인접 픽셀들 간의 상관도가 높다는 특성을 활용하여 저장해야하는 파워 마스크의 크기를 줄이는 방법을 제안한다. 공개된 영상 데이터셋에 다양한 왜곡을 가한 후 핑거프린트 정합 실험을 수행하여, 파워 마스크를 통해서 오인식률을 1/3 이상 줄일 수 있음을 확인하였다.

본 논문은 파워 마스크 기반 영상 핑거프린트 정합에 관한 연구이다. 파워 마스크를 영상 핑거프린팅에 적용하고, 파워 마스크 크기를 줄이는 방법을 제안하였다. 2장에서 영상 핑거프린트 및 파워 마스크 추출 방법을 살펴보고, 파워 마스크 크기를 줄이는 방법을 제안한다. 3장에서 파워 마스크 크기에 따른 핑거프린트 정합 성능 향상 정도를 실험으로 확인하고 분석한다.

2. 파워 마스크를 이용한 영상 핑거프린트 정합

핑거프린트와 핑거프린트의 강인도 정보인 파워 마스크를 추출하는 방법을 살펴보고, 파워 마스크 크기를 줄이는 방법을 제안한다.

2.1 영상 핑거프린트 및 파워 마스크 추출

파워 마스크를 이용한 핑거프린팅 시스템은 Fig. 1에 도시한 바와 같이 먼저 영상들로부터 이진 핑거프린트와 파워 마스크를 추출하여 데이터베이스를 구성한다. 미지의 영상을 식별하기 위해서 핑거프린

트를 이용하여 데이터베이스를 검색할 때 파워 마스크가 정합 가중치로 활용된다. 파워 마스크기반 핑거프린트 정합은 임의의 이진 핑거프린팅 방법에 적용할 수 있으며, 본 논문에서는 기존 웨이블릿 변환 부호 기반 이진 영상 핑거프린트[10]에 파워 마스크 방법을 적용한다. 웨이블릿 변환의 상세계수(detail coefficients)는 0을 중심으로 대칭적이며 독립 항등 정규 분포(independent and identically distributed Gaussian distribution)인 성질이 있다 [11]. 이를 활용하여 Fig. 2에 도시한 바와 같이 영상을 웨이블릿 변환을 통해 분해한 후에 상세계수의 부호를 취하면 0과 1의 분포가 균등한 핑거프린트를 얻게 된다. 먼저 입력 영상 I 를 512×512 로 스케일링한 후에 5계층 Haar 웨이블릿 변환하고, Fig. 2에 나온 바와 같이 상세계수 중에서 대각 영역을 취하여 $N \times M$ 이차원 신호 U 를 얻는다. 다음과 같이 구한 신호 U 의 부호를 취하여 $N \times M$ 이진 핑거프린트 H 를 얻는다.

$$H[n,m] = \begin{cases} 1 & \text{for } U[n,m] > 0 \\ 0 & \text{for } U[n,m] \leq 0 \end{cases} \quad (1)$$

5계층 Haar 웨이블릿 변환의 경우 512×512 영상에 대해서 $N=16, M=16$ 으로 주어지므로 입력 영상에 대해서 총 256비트 핑거프린트가 얻어진다. 일반적으로 두 영상으로부터 각각 얻은 이진 핑거프린트 H_1 과 H_2 간의 정합은 해밍 거리인 D_H 를 사용하며, 개별 비트간 거리를 $d[n,m]=\text{XOR}(H_1[n,m], H_2[n,m])$ 라고 할 때 다음과 같이 주어진다.

$$D_H(H_1, H_2) = \frac{1}{NM} \sum_{n=1}^N \sum_{m=1}^M d[n,m] \quad (2)$$

식(1)의 핑거프린트 추출 과정과 식(2)의 정합 과정에서 U 의 부호만을 취하고 U 의 크기 값을 활용하지 않았다. 따라서 U 의 크기 정보는 손실되며, 이러한 정보 손실을 정합에서 보완하기 위해서 파워 마스크가 제안되었다 [7]. 파워 마스크 방법은 U 의 크기가 클수록 부호를 변경하기 위해서 많은 변형을 가해

야한다고 가정하고, U 의 크기 값을 예측 강인도로 사용한다. 즉, 예측 강인도인 $|U[n,m]|$ 의 값들을 크기순으로 내림차순으로 정렬하고, 크기순으로 큰 T 개의 값을 가지는 비트의 위치를 구한다. 그 T 개 비트 위치의 값을 1로, 나머지 $NM-T$ 개 비트 위치의 값을 0으로 하면 $N \times M$ 이진 파워 마스크 $P[n,m]$ 이 얻어진다. 파워 마스크 $P[n,m]$ 값이 1인 T 비트는 강인한 비트로, $P[n,m]$ 값이 0인 $NM-T$ 비트는 연약한 비트가 된다. 이렇게 구한 파워 마스크 $P[n,m]$ 을 이용하여, 두 핑거프린트 H_1 과 H_2 사이의 가중 해밍거리 D_M 을 강인한 비트와 연약한 비트에 대해서 서로 다른 가중치 α 와 β 를 사용하여 다음과 같이 구하게 된다. 본 논문에서는 기존 논문[7]과 같이 $\alpha=0.5, \beta=1$ 사용하였다.

$$D_M(H_1, H_2) = \frac{\sum_{n=1}^N \sum_{m=1}^M \alpha(1-P[n,m])d[n,m]}{\alpha(NM-T) + \beta T} + \frac{\sum_{n=1}^N \sum_{m=1}^M \beta P[n,m]d[n,m]}{\alpha(NM-T) + \beta T} \quad (3)$$

2.2 파워 마스크 부표분화와 핑거프린트 정합

핑거프린트와 파워 마스크를 사용하면 잡음 등에 더 강인하게 핑거프린트 정합을 수행할 수 있으나, Fig. 1에 도시한 바와 같이 핑거프린트 데이터베이스와 파워 마스크 데이터베이스를 같이 저장해야하므로 저장 공간을 많이 차지하는 단점이 있다. 따라서 영상 식별 서비스에 파워 마스크 개념을 실제로 적용하기 위해서는 핑거프린트와 같은 크기의 저장 공간을 필요로 하는 파워 마스크의 크기를 줄일 필요가 있다. 본 논문에서는 영상들이 공간적 상관도(spatial correlation)가 크다는 성질을 활용하여 파워 마스크의 크기를 줄였다. 즉, 공간 영역에서 인접 픽셀들은 비슷한 값을 가질 확률이 높다. 2.1장에서 Haar 웨이블릿 변환을 통해서 얻은 $U[n,m]$ 은 상세계수의 대각 영역으로 영상의 가로와 세로 방향으로 변화도를 구한 것이므로 영상의 에지 또는 텍스처에서 크기 값이 크게 된다. 또한 영상은 중요한 정보가 들어있는 전경이 있고, 상대적으로 정보량이 적은 배경 부분이 있다. 일반적으로 전경의 예측 강인도가 높게 되며, 파워 마스크 값도 1의 값을 가지게 되고, 배경 부분의 경우 0의 값을 가진다. 즉, 어떤 영상에서 얻은 파워 마스크의 1과 0의 값 분포가 공간적으로 균등하지

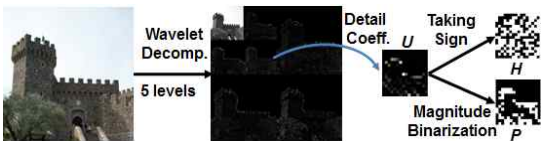


Fig. 2. Image fingerprint and power mask extraction based on the diagonal detail coefficients of the Haar wavelet transform.

않고, 특정 영역(주로 전경 부분)으로 몰리는 현상이 생기게 된다. 따라서 예측 강인도인 $|U[n,m]|$ 도 공간 방향 상관도가 높게 된다. 이러한 파워 마스크 값의 공간 방향 상관도를 이용하면 파워 마스크 크기를 줄일 수 있다.

Fig. 2의 영상 핑거프린트 추출 과정에서 웨이블릿 변환의 $N \times M$ 대각 성분인 U 의 부호를 취하여 핑거프린트 비트로 사용하고, U 의 절대값은 예측 강인도로 사용한다. 예측 강인도 크기 순서에 따라 이진화를 수행하여 파워 마스크를 얻는다. 예측 강인도로 사용하는 대각 성분의 크기 값은 공간 영역 상관도가 높으므로 부표본화(subsampling)를 하더라도 정보의 손실이 크지 않다. 따라서 예측 강인도인 $|U[n,m]|$ 을 가로축과 세로축 모두 $1/R$ 로 부표본화하여 $(N/R) \times (M/R)$ 크기의 예측 강인도를 얻고, 부표본화된 예측 강인도를 크기 값 순서로 나열하고 크기가 큰 T/R^2 개는 파워 마스크 값을 1로, 나머지 위치인 $(NM-T)/R^2$ 개는 파워 마스크 값을 0으로 하여 이진 파워마스크를 구한다. 편의상 본 논문에서 N 과 M 은 R 로 나누어진다고 가정한다. 이렇게 하면 $N \times M$ 파워 마스크 P 를 $(N/R) \times (M/R)$ 파워 마스크 P_R 로 바꾸어 저장 공간을 줄일 수 있다. 본 논문에서는 R 값으로 2, 4, 8의 세 가지 경우를 고려하였고, 얻어진 파워 마스크는 Fig. 3에 도시하였다. Fig. 3에서는 차이점을 쉽게 알 수 있도록 부표본화된 P_R 을 원래의 파워 마스크인 P 크기로 업샘플링(upsampling)하여 도시하였다. 부표본화된 파워 마스크를 사용하면 다음 수식과 같이 $R \times R$ 블록별로 해밍거리를 구하고

해당 블록의 파워 마스크값을 가중치로 하여 가중합을 구하여 두 핑거프린트 H_1 과 H_2 사이의 가중 해밍 거리 D_R 을 구한다.

$$D_R(H_1, H_2) = \frac{\sum_{i=1}^{N/R} \sum_{j=1}^{M/R} \alpha(1 - P_R[i,j]) \sum_{n=1}^R \sum_{m=1}^R d[(i-1)R+n, (j-1)R+m]}{\sum_{i=1}^{N/R} \sum_{j=1}^{M/R} \beta P_R[i,j] \sum_{n=1}^R \sum_{m=1}^R d[(i-1)R+n, (j-1)R+m]} + \frac{\alpha(NM-T) + \beta T}{\alpha(NM-T) + \beta T} \quad (4)$$

식 (3)과 (4)를 구현 시에 파워 마스크 $P[n,m]$ 과 $d[n,m]$ 이 모두 1 또는 0의 값을 가지므로 경우의 수는 4가지가 된다. 이중 $d[n,m]$ 이 0이면 $P[n,m]$ 에 상관없이 0에 대응하고, $d[n,m]$ 이 1인 경우 $P[n,m]$ 이 1이면 β 에 대응하고 0이면 α 에 대응시킨다. 이렇게 핑거프린트 비교 결과를 0, α , β 로 대응시키고 최종적으로 더하기만 수행하도록 하면 곱하기 없이 구현 가능하다.

3. 실험 결과

기존 해밍 거리 기반 영상 핑거프린트 정합 성능과 파워 마스크를 이용한 가중 해밍 거리를 이용한 정합 성능을 실험적으로 비교하였다. 정합 성능 비교를 위해서 실내외, 운송수단, 사람, 풍경 등 다양한 종류의 영상을 포함하는 MIR Flickr-25K 영상 데이터셋 중 5000개의 영상을 사용하였다. 2.1장에서 기술한 바와 같이 원본 실험 영상들을 512×512 로 크기를 조정하고 5계층 Haar 웨이블릿 변환을 수행하고, 상세계수 중에서 대각 영역을 취하여 16×16 크기의 핑거프린트 비트를 얻고($N=16, M=16$), 이 과정에서 파워 마스크도 구한다. 이때 파워 마스크는 부표본화율인 R 값에 따라서 Fig. 3에 도시한 바와 같이 16×16 ($R=1$), 8×8 ($R=2$), 4×4 ($R=4$), 2×2 ($R=8$)의 크기를 가진다. 실험영상에서 추출한 핑거프린트와 파워 마스크로 데이터베이스를 구성하고 핑거프린트 정합 실험을 수행하였다.

일반적으로 콘텐츠 식별 시스템의 성능 비교에는 receiver operating characteristic (ROC) 곡선이 이용된다. ROC 곡선은 인식 시스템에 존재하는 두 가지 형태의 오인식율인 false alarm rate (FAR)와 false rejection rate (FRR)를 가로와 세로축으로 하여 그래프를 그린 것이다. 영상 핑거프린팅 시스템에서 FAR은 서로 다른 영상을 같다고 판정할 확률이며, FRR은 같은 영상을 다르다고 판정할 확률이다.

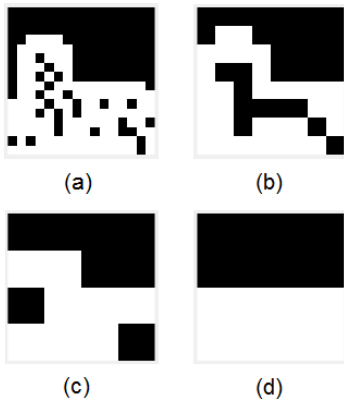


Fig. 3. Power mask of an image: (a) $R=1$ (16×16 , without subsampling) (b) $R=2$ (8×8) (c) $R=4$ (4×4) (d) $R=8$ (2×2).

기존 해밍 거리 D_H , 파워 마스크 기반 거리 D_M , 부표분화된 파워 마스크 기반 거리 D_R 을 공정하게 비교하기 위해서 각 영상으로부터 얻은 16×16 크기의 이진 핑거프린트 (총 256비트)와 파워 마스크를 이용하여 ROC 곡선을 구하였다. FAR을 구하기 위해서는 5000개의 영상으로부터 추출한 핑거프린트와 파워 마스크 데이터베이스를 구축하고, 임의로 2개의 핑거프린트를 선택하고 선택된 핑거프린트 쌍들 간의 거리인 D_H, D_M, D_R 을 각각 구하였다. 영상 식별기의 문턱값을 변화시켜가면서 문턱값 보다 작은 거리를 가지는 핑거프린트 쌍의 비율을 구하였다. FRR을 구하기 위해서는 원본 영상에 왜곡을 가한 후 핑거프린트를 추출하고 원본 핑거프린트와 비교해야한다. 본 논문에서 고려한 영상 변형들은 다음과 같다.

- WN: Adding Gaussian white noise with a mean of zero and a standard deviation of 50.
- FT1: Applying the 5 by 5 median filter.
- FT2: Applying the 3 by 3 average filter.
- FT3: Applying the 3 by 3 unsharp contrast enhancement filter H given by

$$H = \begin{bmatrix} -0.1667 & -0.6667 & -0.1667 \\ -0.6667 & 4.3333 & -0.6667 \\ -0.1667 & -0.6667 & -0.1667 \end{bmatrix}$$
- HQ: Histogram Equalization.
- DT: Dithering with MATLAB default parameters.
- JPG: JPEG compression of the quality factor 75%.

본 논문에서 다루는 웨이블릿 변환 기반 핑거프린트[10]는 기하적인 변형을 고려하지 않은 방법이므로 실험에서 기하적인 변형은 제외하였다. 하지만 제안한 파워 마스크는 기하적인 변형을 고려하는 다른 영상 핑거프린팅 방법[1]에 큰 수정 없이 적용할 수 있다. 영상 변형들을 4개의 집합으로 묶어서, 4종류의 복합 영상 변형을 각 영상에 가하고 핑거프린트를 추출한 후에 원본 영상의 핑거프린트 데이터베이스와 비교하여 D_H, D_M, D_R 을 각각 구하였다. 영상 식별기의 문턱값을 변화시켜가면서 문턱값 보다 큰 거리를 가지는 핑거프린트 쌍의 비율을 구하여 FRR을 구하였다. Fig. 4는 파워 마스크 이진화 문턱값 T 를 바꿔가면서 얻은 ROC 곡선이다. 고려한 모든 영상

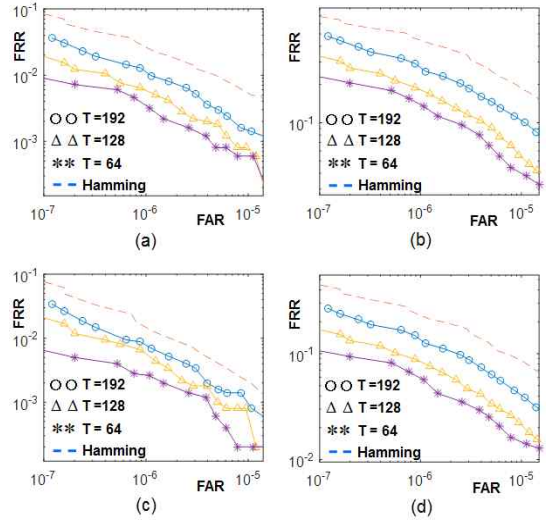


Fig. 4. ROC curves of the Haar-wavelet fingerprint with the weighted Hamming distance based on the power mask of the binarization threshold T for four sets of distortions, (a) FT1+FT2+FT3+JPG2, (b) WN+JPG2, (c) HQ+JPG2, (d) DT+JPG2.

변형들에 대해서 파워 마스크 기반 거리인 D_M 이 기존 해밍 거리 D_H 에 비해서 핑거프린트 정합 성능을 크게 개선함을 알 수 있다. 파워 마스크는 T 개의 강인한 비트와 $NM - T$ 개의 연약한 비트로 구성된다. 이진화 문턱값이 작으면 강인한 비트의 개수가 줄어들어 신뢰도가 높은 소수의 강인한 비트들에 대한 가중치를 높이는 효과가 있다. 반면 이진화 문턱값이 크면 강인한 비트의 개수가 늘어나서 상대적으로 가중치의 효과는 줄어든다. Fig. 4의 결과를 보면 영상의 경우 T 값을 상대적으로 작게 하는 것이 FRR을 줄여서 ROC 곡선 상에서 더 좋은 성능을 보임을 알 수 있다. 하지만 고려한 T 값 모두에서 파워 마스크 기반 거리인 D_M 이 기존 해밍 거리인 D_H 보다 우수한 성능을 보였다. Fig. 5는 파워 마스크의 크기를 가변하면서 ROC 곡선을 구한 것이다. 이진화 문턱값 T 는 64로 고정하였다. 부표분화 정도가 커짐에 따라서 ROC 곡선에서 성능 열화가 확인되지만, 부표분화 정도를 $R=8$ 로 하여 크기를 1/8로 줄인 2×2 크기(즉, 영상 당 4비트)의 이진 파워 마스크를 사용하더라도 제안한 부표분화된 파워 마스크 기반 거리인 D_R 이 기존 해밍 거리인 D_H 보다 우수한 성능을 보였다. 제안한 부표분화를 이용한 파워 마스크 압축이 효과적임을 알 수 있다.

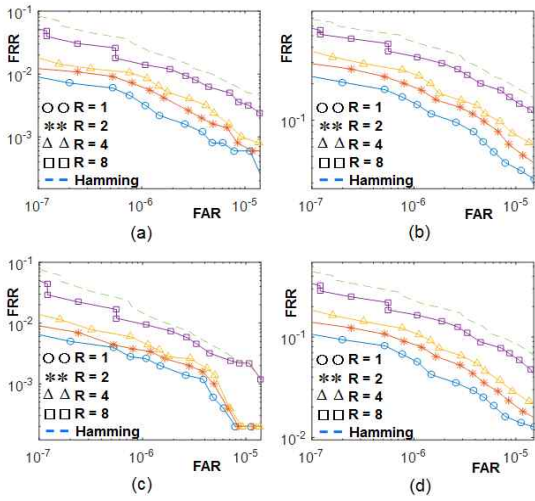


Fig. 5. ROC curves of the Haar-wavelet fingerprint with the weighted Hamming distance based on the subsampled power mask of the subsampling ratio $1/R$ for four sets of distortions. (a) FT1+FT2+FT3+JPG2, (b) WN+JPG2, (c) HQ+JPG2, (d) DT+JPG2.

Table 1은 FAR을 10^{-6} 으로 고정하고, 각 영상변형 별로 FRR을 구하여 정리한 것이다. 성능 비교를 위해서 제안된 파워 마스크 기반 거리의 FRR을 기존 해밍 거리의 FRR로 나눈 relative error rate (RER)을 정의한다. 제안된 방법의 RER이 1이면 성능이 같은 것이며, 1보다 크면 성능이 열화된 것이고, 1보다 작으면 성능이 개선된 것이다. 파워 마스크 기반 거리인 D_M 을 사용할 경우 RER이 0.12에서 0.31 사이의

값을 가졌고 평균값은 0.22였다. 즉, 실험에 사용한 4종류의 복합 영상 변형들에 대해서 기존 D_H 를 사용한 경우와 비교해서 오인식률이 평균 0.22배 수준으로 줄어드는 것을 확인했다. 파워 마스크 방법은 이렇게 오인식률을 크게 줄여주지만, 기존 해밍 거리 기반 정합 방법에서는 256비트의 핑거프린트만 필요한 것에 비해서 파워 마스크 기반 정합 방법에서는 추가로 256비트의 파워 마스크도 저장해야하므로 영상 하나당 데이터베이스에 512비트를 저장해야 한다. 제안된 부표본화된 파워 마스크는 부표본화율 R 을 조정하여 파워 마스크 저장에 필요한 저장 공간을 줄인 것이다. 부표본화율 R 을 크게 하면 파워 마스크 저장을 위해서 필요한 저장 공간은 줄어들지만 정합 성능의 감소도 관찰되었다. 하지만 $R=8$ 을 사용하여 최대 크기를 줄인 파워 마스크를 이용한 D_R 도 D_H 에 대비해서 RER이 0.48에서 0.73 사이의 값을 가졌고, 평균값으로 0.63을 가졌다. 즉 오인식률을 0.63배 수준으로 줄여줄 수 있다. 영상 식별 시스템에 파워 마스크를 실제 적용 시에는 저장 공간과 식별 성능 간의 상충관계 (trade off)를 보고 R 값을 정하게 된다. 파워 마스크 방법은 변형에 강인할 것으로 예상되는 비트들에 대한 가중치를 높이고, 강인하지 않을 것으로 예상되는 비트들에 대한 가중치는 줄인다. 실험 결과에 따라서 2장에서 가설한 영상을 변형시키면 웨이블릿 변환 값이 바뀌게 되는데, 상대적으로 크기 값이 큰 웨이블릿 변환 계수의 부호가 바뀔 확률이 작다는 것이 검증되었음을 알 수 있다.

Table 1. FRR (%) of the fingerprint matching for FAR= 10^{-6}

Matching method		Image distortions			
		FT1+FT2+FT3+JPG2	WN+JPG2	HQ+JPG2	DT+JPG2
Hamming D_H (256 bits)	FRR	2.90	45.56	1.52	23.78
	RER	1	1	1	1
Weighted Hamming D_M with $T=64, R=1$ (512 bits)	FRR	0.34	14.08	0.28	6.12
	RER	0.12	0.31	0.18	0.26
Weighted Hamming D_R with $T=64, R=2$ (320 bits)	FRR	0.64	19.40	0.36	8.70
	RER	0.22	0.43	0.24	0.37
Weighted Hamming D_R with $T=64, R=8$ (260 bits)	FRR	1.40	33.38	0.94	16.46
	RER	0.48	0.73	0.62	0.69

The number in the parentheses is the number of bits needed for storing fingerprint and power mask for each image. RER refers to the relative error rate, which is defined by the FRR of the proposed weighted distance divided by the FRR of the pervious Hamming distance.

4. 결 론

오디오 핑거프린팅에 사용되었던 파워 마스크 방법을 영상 핑거프린팅에 적용하여 오인식을 개선 정도를 확인하였다. 파워 마스크는 영상에서 추출한 핑거프린트 비트의 예측 강인도를 이진화한 것이다. 파워 마스크를 사용하면 고려한 모든 변형들에 대해서 오인식률이 1/3 이상 줄어드는 것을 확인하였다. 파워 마스크는 핑거프린트 추출 과정에서 구해지며, 이진화하여 핑거프린트와 같이 데이터베이스에 저장한다. 본 논문에서는 파워 마스크를 위한 저장 공간을 줄이기 위해서, 영상의 공간적 상관도를 이용하여 부표본화된 파워 마스크를 추출하는 방법을 제안하였다. 다양한 영상 변형들에 대한 실험을 통해서 파워 마스크가 핑거프린트의 정합 성능을 제고함을 보였고, 부표본화된 파워 마스크가 데이터베이스 저장에 필요한 공간을 줄이는데 효과적임을 확인하였다.

REFERENCE

[1] J. Seo, T. Kalker, J. Haitsma, and C. Yoo, "A Robust Image Fingerprinting System Using the Radon Transform," *Signal Processing: Image Communication*, Vol. 19, No. 4, pp. 325-339, 2004.

[2] X. Nie, X. Li, Y. Chai, C. Cui, X. Xi, Y. Yin, et. al., "Robust Image Fingerprinting Based on Feature Point Relationship Mining," *IEEE Transactions on Information Forensics and Security*, Vol. 13, No. 6, pp. 1509-1523, 2018.

[3] M. Kim, M. Park, and H. Kim, "Automatic Music Summarization Method by Using the Bit Error Rate of the Audio Fingerprint and a System Thereof," *Journal of Korea Multimedia Society*, Vol. 16, No. 4, pp. 453-463, 2013.

[4] R. Venkatesan, S.M. Koon, M.H. Jakubowski, and P. Moulin, "Robust Image Hashing," *Proceeding of International Conference on Image Processing*, pp. 664-666, 2000.

[5] Y. Li and L. Guo, "Robust Image Fingerprinting via Distortion-resistant Sparse Coding," *IEEE Signal Processing Letters*, Vol. 25, No. 1, pp. 140-144, 2018.

[6] Z. Tang, L. Chen, X. Zhang, and S. Zhang, "Robust Image Hashing with Tensor Decomposition," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 31, No. 3, pp. 549-560, 2019.

[7] B. Coover and J. Han, "A Power Mask Based Audio Fingerprint," *Proceeding of International Conference on Acoustic, Speech and Signal Processing*, pp. 1394-1398, 2014.

[8] J. Seo, "An Asymmetric Matching Method for a Robust Binary Audio Fingerprinting," *IEEE Signal Processing Letters*, Vol. 21, No. 7, pp. 844-847, 2014.

[9] J. Seo, "Perceptual Bound-based Asymmetric Image Hash Matching Method," *Journal of Korea Multimedia Society*, Vol. 20, No. 10, pp. 1619-1627, 2017.

[10] A. Varna, A. Swaminathan, and M. Wu, "A Decision Theoretic Framework for Analyzing Binary Hash-based Content Identification Systems," *Proceeding of ACM Workshop on Digital Rights Management*, pp. 67-76, 2008.

[11] S. Mallat, "A Theory for Multiresolution Signal Decomposition: the Wavelet Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 7, pp. 674-693, 1989.



서진수

1998년 한국과학기술원 전기 및 전자공학과 공학사
 2000년 한국과학기술원 전기 및 전자공학과 공학석사
 2005년 한국과학기술원 전기 및 전자공학과 공학박사

2006년~2008년 한국전자통신연구원 선임연구원
 2008년~현재 강릉원주대학교 전자공학과 교수