creative commons

COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

공학석사 학위논문

핑거프린팅 기법 및 영상 처리를 활용한 레코드 정보 검색 연구

Vinyl Record Information Retrieval via Audio Fingerprinting and Image Processing

2015 년 8 월

서울대학교 융합과학기술대학원

융합과학부 디지털정보융합전공

원 상 희

국문초록

디지털 기술의 발달로 대부분의 음악 소비자가 디지털 음원을 감상하는 시대가 되었다. 또한 통신 기술이 발달함에 따라 개인 기기에 음원을 저장하지 않고 스트리밍(Streaming) 서비스를 활용해 음악을 듣는 사용자가 늘어나고 있다. 하지만 이러한 흐름 속에서 역설적으로 아날로그 매체인 LP 레코드의 수요는 빠르게 증가하고 있다.

디지털 음원과 달리 아날로그 음원에는 메타데이터를 저장할 수 없고, 따라서 음원과 관련된 정보를 저장 및 시각화하는 것도 불가능하다. 디지털 음원에 저장되는 메타데이터는 곡의 제목, 아티스트, 조성, 빠르기, 실시간 진행 등의 정보를 포함하고, 이는 기기에서 시각화 됨으로써 일반 사용자에게 편의를 제공할 뿐 아니라, DJ의 공연에 있어 요긴하게 활용된다.

본 연구에서는 오디오 핑거프린팅 기법 및 영상 처리를 활용하여 LP 레코드에서도 메타데이터를 실시간 활용할 수 있도록 하는 시스템을 제안한다. 제안하는 시스템은 오디오 핑거프린팅 기법을 활용하여 LP 레코드를 식별하고, 오디오 신호 처리를 통해 재생 중인 음원을 실시간 정렬함으로써 사용자에게 음원의 정보를 제공한다. 하지만 같은 패턴의 반복이 많은 현대 음악의 특성상 오디오 신호 처리 만으로는 실시간 정렬에 오류가 발생한다. 이를 보정하기 위해 LP 레코드 이미지의 회전각을 영상 처리를 통하여 추정하였고, 그 성능을 평가하기 위해 기존의 오디오 상호 상관 만을 사용하였을 때와 비교하였다.

주요어 : 뮤직 테크놀로지, 레코드, 음악 정보 검색, 실시간 정렬, DJ

학번 : 2013-22409

목 차

제 1 장 서론 1
제 1 절 연구 배경 1
제 2 절 논문의 구성 6
제 2 장 이론적 배경 7
제 1 절 오디오 핑거프린팅 7
제 2 절 오디오 신호 정렬 12
제 3 절 SURF: Speeded Up Robust Features 15
제 3 장 연구 방법 및 제안 시스템 17
제 1 절 연구 방법 17
3.1.1. 오디오 핑거프린팅 17
3.1.2. 프로토타입 실험 17
3.1.3. 입력 신호 정렬 20
제 2 절 제안 시스템 21
제 4 장 실험 결과 23
제 1 절 오디오 핑거프린팅 23
제 2 절 입력 신호 정렬 23
4.2.1. SURF 이미지 23
4.2.2. 프로토타입 실험 25
4.2.3. 디지털 음원과의 비교 29
4.2.4. 입력 신호 정렬 31
제 5 장 결론 35
참고문헌 37
Abstract 39

그림 목차

그림 1 음원 정보를 제공하는 인터페이스 3

그림 2 DJ 공연에 활용되는 인터페이스 3

그림 3 3초의 길이를 가지는 음원의 오디오 핑거프린트 예시. 4

그림 4 반복이 많은 곡의 자기 상관 행렬 5

그림 5 오디오 핑거프린팅 시스템의 구조 7

그림 6 스펙트로그램 10

그림 7 스펙트로그램 상의 유의미한 지점 10

그림 8 오디오 신호 정렬에 오류가 발생하는 예 14

그림 9 회전하는 레코드의 SURF 관심점 15

그림 10 프로토타입 실험 환경 19

그림 11 본 실험 환경.. 19

그림 12 실시간 정렬을 위한 제안 시스템.. 21

그림 13 충분한 SURF 관심점 24

그림 14 불충분한 SURF 관심점.. 24

그림 15 제안 시스템에 활용될 수 없는 예.. 24

그림 16 안정적으로 변화하는 LP 레코드 위상.. 26

그림 17 오차가 존재하는 LP 레코드 위상.. 27

그림 18 70초-73초 구간의 입력 신호를 넣었을 때의 오디오 상호 상관.. 28

그림 19 디지털 음원과 비교했을 때의 상호 상관과 아날로그 음원과 비교했을 때의 상호 상관 30

그림 20 안정적으로 변화하는 위상 (60 fps).. 31

그림 21 오디오 상호 상관이 높은 55초 구간에서의 위상.. 33

그림 21 오디오 상호 상관이 높은 67초 구간에서의 위상.. 33

그림 21 SURF 관심점이 잘못 연결된 예.. 34

표 목차

표 1 프로토타입 실험에 사용된 LP 레코드 18

표 2 본 실험에 사용된 LP 레코드 20

표 3 오디오 상호 상관을 사용했을 때의 정렬 정확도와 프로토타입 실험에서의 정렬 정확도 비교 29

표 4 오디오 상호 상관을 사용했을 때의 정렬 정확도와 제안 시스템의 정렬 정확도 비교 32

제 1 장 서론

제 1 절 연구 배경

기술의 발달과 더불어 음악을 저장하는 매체의 종류도 꾸준히 변해왔다. 특히 디지털 매체의 등장과 통신 기술의 발달은 음악 소비 행태를 크게 바꿔놓았다. 음반을 소유하던 시대에서 음원을 다운로드 하는 시대로 변화하였고, 자신의 기기에 음원을 저장하지 않고도 스트리밍 서비스(Streaming service)를 이용해 음악을 듣는 시대가 되었다. CD(Compact Disc)의 판매량은 감소하고 있고, 2013년에는 아이튠즈 뮤직 스토어가 문을 연 2003년 이래 처음으로 디지털 음원 판매량이 감소했다고 한다[1]. 하지만 이와 반대로 아날로그 매체인 LP(Long Play) 레코드의 판매량은 다시 큰 폭으로 성장하고 있다.

1980년대 초반, CD라는 디지털 매체가 등장한 이후 LP 레코드 시장은 사양길로 접어들었다. 기존의 LP 레코드에 비해 부피도 작고 반영구적인 이 매체는 대중들에게 큰 환영을 받았고, 휴대용 CD 플레이어의 등장으로 휴대성까지 지니게 되었다. 또한 LP 레코드의 약점이었던 저역대의 정보도 충실히 담아낼 수 있고, 잡음이 없는 깔끔한 소리를 재생할 수 있었다. 무엇보다 제작 단가가 LP 레코드의 1/5 정도 임에도 판매 가격은 더 높았기에 제작자들 역시 LP 레코드 보다는 CD 제작을 선호하게 되었다. 이렇듯 그 역할이 대체되어 판매가 급감했던 LP 레코드의 판매량은 2006년 이후 다시 서서히 증가하기 시작했고, 물리적 음반을 구매하지 않는 시대적 흐름 속에서도 그 성장폭이 점점 커지고 있다. 미국의 인터넷 종합 쇼핑몰 아마존의 경우, 가장 빠른 성장세를 보이는 음악 매체가 LP 레코드로, 5년전 대비 745%에 이르는

성장률을 보였다고 한다[2]. LP 레코드를 경험한 세대 뿐 아니라 이를 경험하지 않은 젊은 세대들 또한 소비층을 구성한다. 과거에 대한 향수, 물리적 음반으로서의 매력, 빠른 것만을 추구하는 현대 사회에 속에서 여유를 찾고자 하는 움직임 등 다양한 동기로 소비자는 LP 레코드를 구입한다. 일반 사용자 외에도 일부 DJ(Disc Jockey) 들은 자신들의 공연을 위해 이 아날로그 매체들을 수집한다. 디지털 음원에 비해 불편하고 휴대성이 떨어짐에도 불구하고 전통적 방식을 고수하는 DJ들은 여전히 LP 레코드를 사용한다[3].

CD 플레이어나 mp3플레이어, 스마트폰 등을 통해 디지털 음원을 감상할 때, 재생 기기에서는 곡의 제목, 아티스트, 실시간 진행 등의 정보를 시각화 하여 사용자에게 제공한다 (그림 1). 또한 각 곡의 음악적 특성들은 DJ의 공연에 유용하게 활용되기에 곡의 조성이나 빠르기, 믹싱 구간 등의 정보를 추가로 저장하여 이를 활용하기도 한다 (그림 2). 음악을 끊기지 않고 자연스럽게 이어나가야 하는 DJ들은 시각적으로 제공되는 시간 정보를 활용해 믹싱을 위한 순간을 쉽게 결정할 수 있고, 조성과 빠르기 등의 음악적 정보는 그들의 선곡과 믹싱에 영향을 미친다. 하지만 아날로그 매체는 이러한 메타데이터를 저장할 수 없기 때문에 재생 중인 곡의 정보를 실시간으로 얻을 수 없다. 현재 재생 중인 곡의 정보를 알기 위해선 LP 레코드의 겉면에 적혀 있는 곡 순서를 확인하거나 턴테이블 카트리지의 대략적 위치를 통해 실시간 진행을 짐작하는 수 밖에 없다.

그림 1. 음원 정보를 제공하는 인터페이스

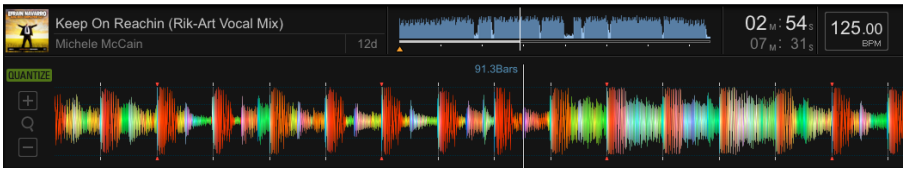

그림 2. DJ 공연에 활용되는 인터페이스

현존하는 오디오 핑거프린팅[4][5][6][7] 기술을 활용하면 짧은 시간 안에 재생 중인 음원을 식별해낼 수 있기에 LP 레코드의 정보를 데이터베이스에 저장하고 불러와 사용자에게 시각화 해 줄 수 있다. 오디오 핑거프린팅 기술은 각 곡의 주파수 대역별 특성을 사람의 지문과 같이 고유의 패턴으로 저장하고 찾아내는 기술이다 (그림 3). LP 레코드에서 나오는 오디오 신호로부터 오디오 핑거프린트를 추출하여 재생 중인 곡을 식별해내면, 미리 구축해 둔 데이터베이스로부터 해당 음원의 정보를 받아올 수 있게 된다.

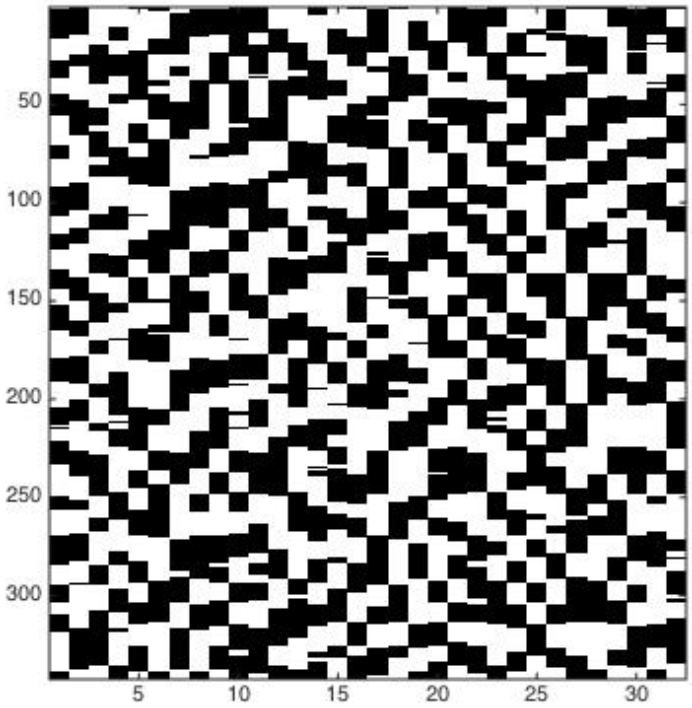

그림 3. 3초의 길이를 가지는 음원의 오디오 핑거프린트 예시

곡의 실시간 진행 정보를 알기 위해서는 추가적으로 오디오 신호를 정렬하는 알고리즘이 필요하다. 하지만 그림 4의 자기 상관 행렬에서 나타난 것과 같이 동일한 패턴의 반복이 많은 현대 음악의 특성으로 인해 오디오 입력 신호 만으로는 현재 재생 중인 구간의 정확한 시간을 파악하지 못하는 경우가 발생할 수 있다. 특히 빠른 정렬을 위해선 보다 짧은 오디오 신호를 사용해야 하는데, 이 경우, 잘못된 지점에 정렬될 가능성은 더욱 높아진다. 이를 극복하기 위해 본 연구에서는 영상 정보를 동시에 활용하려 한다. 아날로그 매체인 LP 레코드는 물리적으로 회전을 함으로써 저장된 오디오 정보를 읽어 온다. 일정한 각속도로 회전하는 LP 레코드 겉면의 이미지를 카메라로 촬영한 뒤, 그 영상 정보를 오디오 신호와 정렬하여 활용하면 오디오 신호 만을 활용했을 때의 잘못된 정렬을 피할 수 있다.

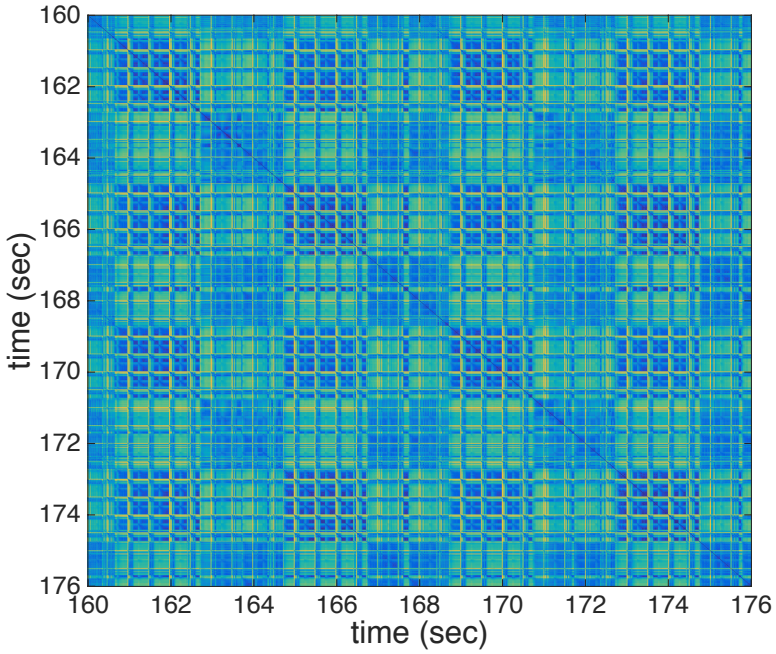

그림 4. 반복이 많은 곡의 자기 상관 행렬

제 2 절 논문의 구성

본 논문에서는 메타데이터를 저장할 수 없는 아날로그 매체인 LP 레코드에서도 실시간 음악 정보 시각화가 가능하도록, 오디오 핑거프린트와 이미지 회전 정보를 활용하여 LP 레코드를 식별 및 정렬하는 시스템을 제안한다. 제2장에서는 제안하는 시스템에서 사용하는 오디오 핑거프린팅 기법과 오디오 정렬 알고리즘, 그리고 영상 처리 알고리즘인 SURF(Speeded-Up Robust Features)에 관하여 서술하였다. 이를 토대로 제3장에서는 제안하는 시스템의 구조에 대하여 자세히 서술하였고, 제4장에서는 시스템을 활용하여 진행한 실험 결과를 기존의 오디오 정렬 알고리즘과 비교, 분석 하였다. 끝으로 제5장에서는 연구의 결론을 도출하고 시스템의 개선 방안 및 연구 방향에 대해 논함으로써 논문을 마무리 지었다.

제 2 장 이론적 배경

제 1 절 오디오 핑거프린팅

과거에는 텔레비전이나 라디오에 모르는 음악이 나오는 경우, 주변 사람들에게 물어보는 것 외에는 그 음악의 정보에 접근할 수 있는 방법이 없었다. 하지만 현재는 스마트폰 어플리케이션을 실행시키는 것만으로도 간편하게 자신이 알지 못하는 음악의 정보를 얻을 수 있다. 이런 어플리케이션에 널리 활용되는 방법 중 하나가 오디오 핑거프린팅(Audio fingerprinting) 혹은 CBID(Content-Based audio Identification) 라고 불리는 기술이다.

핑거프린팅 기술은 실제로 오디오 뿐 아니라 여러 멀티미디어 분야에서 연구되고 있다. 그 중 오디오 핑거프린팅은 오디오 신호로부터 주파수 대역별 특성을 추출하고 저장하는 방식을 의미한다. 앞선 상황에서와 같이 확인되지 않은 오디오로부터 주파수 대역별 특성인 오디오 핑거프린트를 추출하고 이를 오디오 핑거프린트 데이터베이스와 비교하여 일치하는 곡의 정보를 사용자에게 제공해준다 (그림 5).

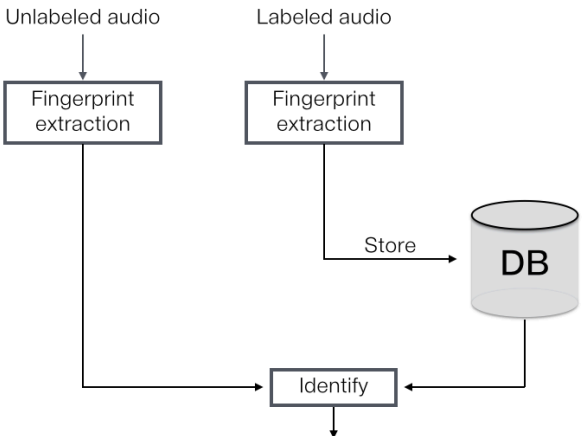

그림 5. 오디오 핑거프린팅 시스템의 구조

알려지지 않은 오디오를 식별하는데 있어 오디오 핑거프린팅과 많이 사용되는 또다른 기술 중 하나가 오디오 워터마킹(audio watermarking)[8][9] 이다. 오디오 워터마킹이란 오디오 파일에서 인간 청각 시스템의 인지를 방해하지 않는 선에서 식별 정보를 저장하고 이를 불러오는 방식으로 주로 저작권 보호를 위해 많이 사용되는 기술이다.

$$y(n) = x(n) + w(n) \tag{2.1}$$

수식 2.1에서와 같이 오디오 신호 $x(n)$에 워터마크 $w(n)$을 삽입해두고, 검출할 때에는 알려지지 않은 신호 $y(n)$과 $w(n)$의 상호 상관을 활용해 오디오 신호를 식별해낸다. 여기서 결과를 결정하는 값 d는 수식 2.2과 같이 나타낼 수 있다.

$$\begin{aligned} y(n) &= x(n) + w(n)d = d_x + d_w \\ &= \frac{1}{N}\sum_i x(i)w(i) + \frac{1}{N}\sum_i w^2(i) \end{aligned} \tag{2.2}$$

삽입되는 워터마크 $w(n)$은 다양한 주파수 대역에 퍼뜨려 두기 때문에 워터마크를 제거하기 위해서는 오디오 파일이 손상될 수 밖에 없다. 하지만 미리 워터마크를 만들어두지 않으면 사용할 수 없다는 단점이 있다. 오디오 핑거프린팅의 경우, 오디오 신호 자체의 특성을 활용하기 때문에 신호 자체를 변형하지 않는 이상 일정한 오디오 핑거프린트를 지니고, 사전 워터마크 삽입 과정이 불필요하기에 제안하는 시스템에 더 적합하다.

[4]에서는 입력된 오디오 신호로부터 매 11.6ms 마다 32비트의 부

분 오디오 핑거프린트(sub-fingerprint)를 추출하였다. 부분 오디오 핑거프린트를 추출하기 위해서는 인간의 인지에서 중요한 주파수 대역인 300Hz에서 2000Hz까지를 바크 비율(Bark scale) 이라 불리는 대수의 대역으로 33등분 한다. n번째 프레임에서 대역 m의 에너지를 E(n,m)이라 하고, n번째 프레임의 m비트 부분 오디오 핑거프린트를 F(n,m)이라 하면 오디오 핑거프린트를 구하는 식은 수식2.3과 같이 나타낼 수 있다.

$$F(n,m)=\begin{cases}1 & if \quad E(n,m)-E(n,m+1)-(E(n-1,m)-E(n-1,m+1))>0 \\ 0 & if \quad E(n,m)-E(n,m+1)-(E(n-1,m)-E(n-1,m+1))\le 0\end{cases} \tag{2.3}$$

추출된 오디오 핑거프린트는 데이터베이스에 저장되고, 새로운 오디오 입력 신호가 들어왔을 때, 입력 신호의 오디오 핑거프린트를 데이터베이스에 저장된 오디오 핑거프린트와 비교하여 일치하는 곡을 검색하여 사용자에게 정보를 제공한다.

$$\mathbf{STFT}\{x[n]\}(m,\omega) = X(m,\omega) = \sum_{n=-\infty}^{\infty} x[n]w[n-m]e^{-j\omega n} \tag{2.4}$$

[5]는 스펙트로그램(Spectrogram) 상에서 유의미한 지점을 찾고, 이를 성좌와 같이 연결하여 오디오 핑거프린트를 추출하였다. STFT(Short Time Fourier Transform, 수식 2.4)를 통해 시간에 따른 주파수 대역의 에너지 분포를 볼 수 있는 스펙트로그램(그림 6)의 특성은 각 오디오의 고유한 특성을 반영하고, 여기서 유의미한 지점(그림 7) 간의 연결을 통해 생성된 오디오 핑거프린트는 오디오를 식별하는 데에 있어 강력한 특징으로 활용될 수 있다.

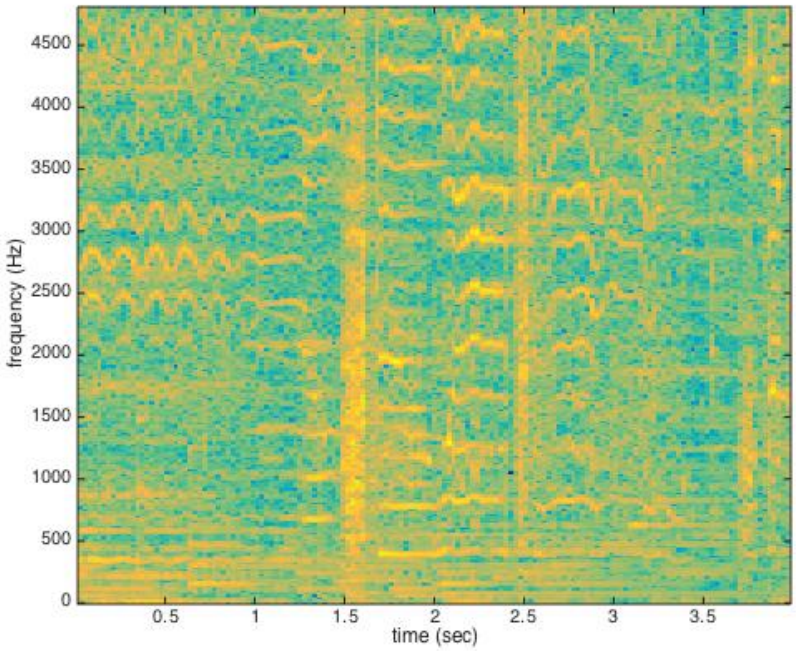

그림 6. 스펙트로그램

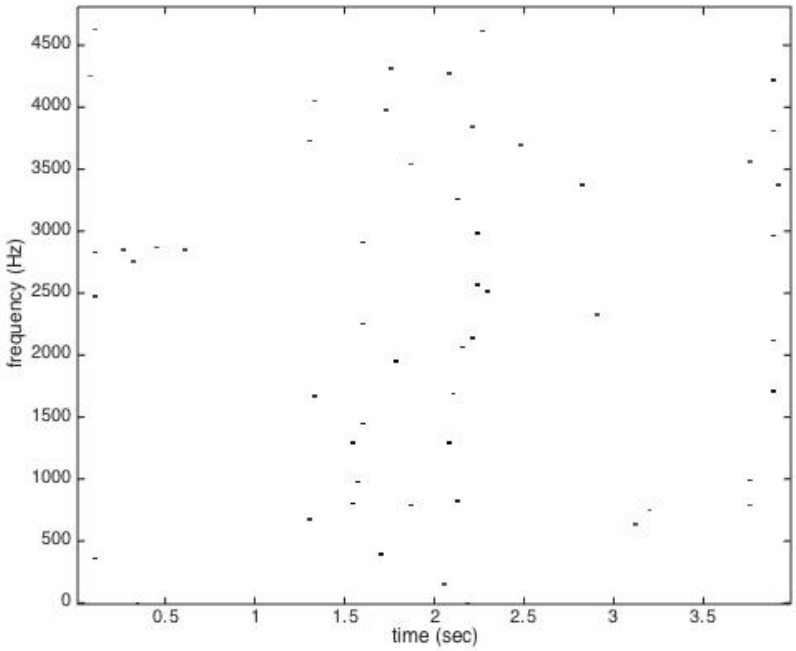

그림 7. 스펙트로그램 상의 유의미한 지점

언급한 다양한 오디오 핑거프린팅 시스템을 평가하기 위해 *"Request for Information on AudioFingerprinting Technologies"*에서 IFPI(International Federation of the Phonographic Industry)와 RIAA(Recording Industry Association of America)는 평가 기준을 다음과 같이 마련하였다[10].

정확도(Accuracy) : 정확한 식별, 실패한 식별, 잘못된 식별(긍정오류).

신뢰도(Reliability) : 저작권 단속을 위한 경우에 가장 중요한 것은 현재 콘텐츠가 재생중인지 아닌지를 판단하는 것이다. 이 경우, 재생되지 않은 음악은 검출이 되어선 안된다.

강인함(Robustness) : 오디오의 압축률, 잡음, 송수신 채널에서의 영향 등에 관계없이 정확히 식별할 수 있는 능력.

입도(Granularity) : 수초 가량으로 짧은 시간의 오디오로부터 전체 오디오를 식별할 수 있는 능력.

보안성(Security) : 조작과 변형을 통해 핑거프린트 검출을 피하는 행위를 막는 능력.

융통성(Versatility) : 오디오 형식에 관계없이 식별할 수 있는 능력. 다른 어플리케이션에서 동일한 데이터베이스를 적용할 수 있는 능력.

확장성(Scalability) : 방대한 데이터베이스 속에서의 성능 혹은 많은 수의 동시 식별 능력.

복잡도(Complexity) : 핑거프린트 추출에 소비되는 계산 비용, 검색

의 복잡도, 핑거프린트 비교의 복잡도, 새로운 정보를 데이터베이스에 기록하는 데 필요한 비용 등.

취성(Fragility) : 일부 어플리케이션에서는 콘텐츠의 변화를 발견할 수 있어야 한다. 이는 강인함과 반대되는 개념으로, 콘텐츠 가공을 위한 변형에는 강인하고 외부 잡음에는 아니어야 한다.

제 2 절 오디오 신호 정렬

오디오 핑거프린팅을 통해 입력된 오디오 신호를 식별한 후에 실시간 재생 구간을 알기 위해서는 오디오 신호 정렬이 필요하다. 오디오 신호 정렬을 위한 기술로는 [11-16] 등이 있다. 이러한 기술들은 악보 추적, 1인 연주자를 위한 자동 반주 생성, 라이브 추적, 실시간 가사 동기화 등 폭넓은 분야에 사용되고 있다.

시간에 따라 변화하는 두 신호의 정렬을 위해 많이 사용되는 방식으로는 동적 시간 워핑(Dynamic Time Warping)[16]을 들 수 있다. 한 연구에서는 동적 시간 워핑을 시간에 따른 각 주파수 대역의 미분값에 적용하여 정렬하는 시스템을 만들었다. STFT를 통해 스펙트로그램을 계산하고, 각 대역을 앞선 오디오 핑거프린팅에서과 같이 대수의 대역으로 나눴다. 각 대역에서의 양수의 미분값 만을 취한 후 동적 시간 워핑을 통해 두 신호를 정렬하였다. [13]에서는 두 신호의 크로마그램(Chromagram)을 추출하여 유클리디안 거리 (Euclidean distance)를 계산하고 여기에 동적 시간 워핑을 적용하였다.

[12]의 연구에서는 MBSES(Multi-Band Entropy Signature)라는 독자적인 오디오 핑거프린트를 활용하여 두 신호를 정렬하려 하였다.

이러한 선행 연구들이 존재하지만 오디오 신호만을 활용한 정렬을 본 시스템에 적용하기에는 어쩔 수 없는 한계가 존재한다. 현대 음악은 비슷한 패턴의 반복이 많고, 1절과 2절에서 유사한 구조가 존재하는 경우가 많다. 또한 컴퓨터 음악이 발달함에 따라, 같은 음악적 패턴은 사람의 연주와 달리 정확히 동일한 신호로 나타나게 된다. 더욱이 하우스나 테크노와 같은 장르의 음악은 전자 음악 중에서도 동일한 패턴의 반복이 많은 경향을 띄기 때문에 오디오 정렬을 하는데에 어려움이 따르게 된다. 같은 음악적 패턴은 같은 스펙트로그램, 같은 크로마그램, 같은 오디오 핑거프린트를 보이므로 오디오 신호만을 활용해서는 정확한 실시간 정렬을 할 수 없다.

본 연구에서는 오디오 신호만을 사용하였을 때의 실시간 정렬이 지니는 한계를 증명하기 위해, 스펙트로그램이나 오디오 핑거프린트와 같은 압축된 형태의 정보가 아닌, 두 오디오 신호 자체의 상호 상관(Cross correlation)을 계산하여 실시간 정렬의 후보군이 다수 존재함을 보이고, 이를 제안하는 시스템을 통해 개선하는 실험을 진행하였다. 상호 상관은 시간에 따른 두 신호의 유사도와 딜레이를 관찰하는데 사용되는 방법으로 수식 2.5와 같이 나타낼 수 있다.

$$(f \star g)(\tau) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f^*(t) g(t+\tau) dt \qquad (2.5)$$

$f(t)$와 $g(t)$는 각기 다른 두 신호를 의미하고 $f^*(t)$는 $f(t)$의 켤레 복소수를 의미한다.

그림 8은 비슷한 패턴의 반복으로 구성된 곡에서 11초부터 14초까지의 3초를 잘라내어 원래 신호와의 상호 상관을 계산한 결과를 나타내고 있다. 이 구간은 곡의 도입부로 반복적인 리듬 패턴을 보이는 구간이다. 그 결과, 11초에서의 상호 상관이 높게 나타나고 있지만 339초와 355.2초에서의 상호 상관이 각각 더 높게 나타나 잘못 정렬될 가능성이 있다. 339초와 355.2초 구간은 곡의 후반부로 도입부와 동일한 리듬 패턴을 보이는 구간이기 때문에 상호 상관이 높게 나타났다.

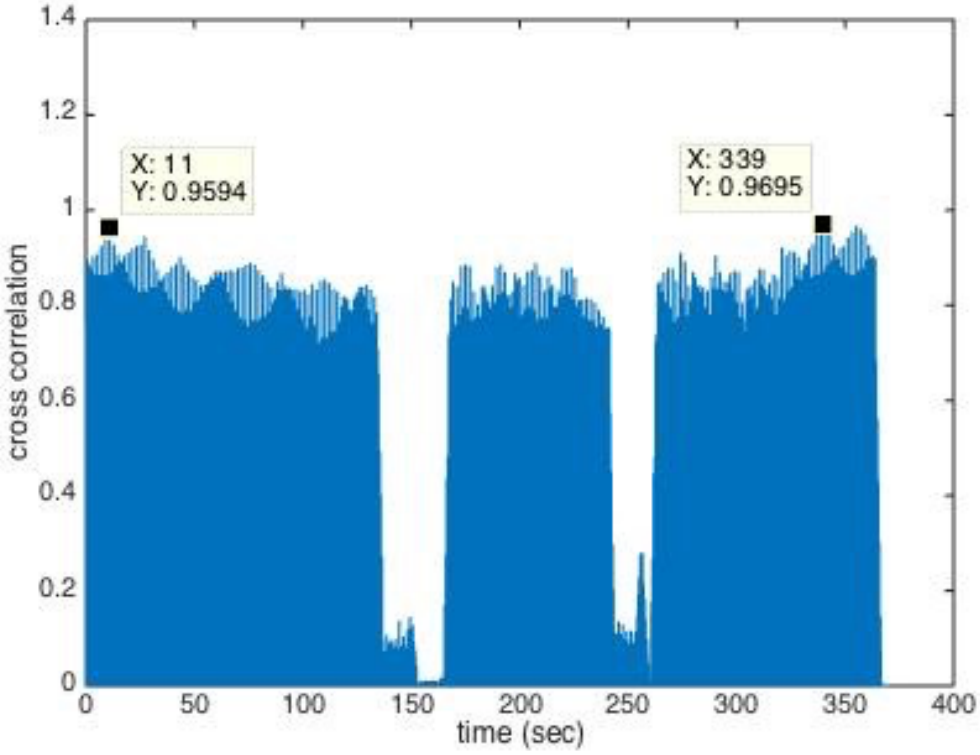

그림 8. 오디오 신호 정렬에 오류가 발생하는 예

제 3 절 SURF: Speeded Up Robust Features

SURF(Speeded Up Robust Features)[17]는 이미지의 크기, 조명, 회전, 시점 등의 환경 변화를 고려하여 이미지의 특징을 찾는 알고리즘이다. SURF는 적분 이미지(integral imgae)를 사용하고, 이미지로부터 관심점(interest point)를 탐지하기 위하여 헤시안 행렬(Hessian matrix)를 사용함으로써 속도와 정확성을 높였다. 이미지 I에서의 한 점 $\mathbf{x}$ = (x,y)에서 크기(scale)가 σ일 때의 헤시안 행렬 $\mathcal{H}(\mathbf{x},\sigma)$는 수식 3.1과 같이 표현할 수 있다.

$$\mathcal{H}(\mathbf{x},\sigma) = \begin{bmatrix} L_{xx}(\mathbf{x},\sigma) & L_{xy}(\mathbf{x},\sigma) \\ L_{xy}(\mathbf{x},\sigma) & L_{yy}(\mathbf{x},\sigma) \end{bmatrix} \tag{3.1}$$

이 때, $L_{xx}(\mathbf{x},\sigma)$는 점 $\mathbf{x}$에서의 x에 대한 가우시안 2차 편미분과 이미지 I의 콘볼루션(convolution)을 의미하고, $L_{xy}(\mathbf{x},\sigma)$와 $L_{yy}(\mathbf{x},\sigma)$는 각각 x, y에 대한 가우시안 편미분과 이미지 I의 콘볼루션, y에 대한 가우시안 2차 편미분과 이미지 I의 콘볼루션을 의미한다.

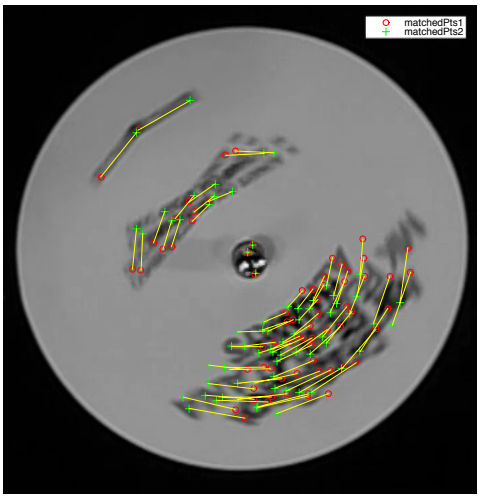

그림 9. 회전하는 레코드의 SURF 관심점

본 연구에서는 LP 레코드 겉면 이미지의 회전 각도를 구하고자 했고, 카메라의 각도나 거리, 조명의 영향 등에서 오는 왜곡을 보완하기 위해 SURF를 사용하였다. 그림 9는 회전하는 LP 레코드 겉면의 이미지로부터 SURF 관심점을 추출하고 각 관심점을 연결하여 회전량을 나타낸 그림이다. O로 표시된 부분은 시작 이미지의 관심점 이고, +로 표시된 부분은 비교 이미지의 관심점이며, 이를 선으로 연결하였다.

제 3 장 연구 방법 및 제안 시스템

제 1 절 연구 방법

3. 1. 1 오디오 핑거프린팅

본 연구에서는 [4]에서 제안한 방식의 오디오 핑거프린팅 기술을 활용해 50곡의 LP 레코드 음원으로부터 오디오 핑거프린트를 추출하고 이를 찾는 성능을 평가하였다. 기존의 연구에서와 동일하게 5000Hz로 다운샘플링 한 음원을 STFT 하였다. 이 때의 윈도우는 31/32의 오버랩 수치를 가지는 해닝 윈도우(Hanning window)를 사용하여 11.6ms 간격으로 부분 오디오 핑거프린트를 추출할 수 있게 설정했다. 각 주파수 대역 역시 기존 연구에서 사용된 33개의 대수 대역으로 설정하였고, 3초의 입력 신호에 대한 성능 평가를 진행하였다.

오디오 신호를 받는 과정에서 Technics 사의 SL-1200 MK5 턴테이블과 Ortofon 사의 Elektro 카트리지를 사용하였고, 이는 Allen & Heath 사의 Xone 92 믹서를 통해 증폭되어 RME 사의 오디오 인터페이스인 Babyface를 통해 48000Hz의 샘플링 레이트를 가지는 디지털 오디오로 변환되어 컴퓨터에 저장되었다.

3. 1. 2 프로토타입 실험

오디오 신호만을 활용한 두 신호의 정렬과 제안 시스템의 성능 비교를 위해 각각 동일하게 두 신호의 오디오 상호 상관을 계산했다. 다섯 장의 LP 레코드에 대해 3초의 길이를 가지도록 신호를 잘라내어 상

호 상관의 입력 신호로 활용했다. 이 때, 실제 사용 상황을 고려하기 위하여 각각의 LP 레코드를 두 번씩 녹음하였고, 새로 녹음한 신호로부터 입력 신호를 잘라낸 뒤, 처음 녹음한 신호와의 상호 상관을 계산하였다. 표 1은 사용된 LP 레코드에 수록된 곡의 제목, 아티스트, 그리고 출시 년도를 나타낸다. 프로토타입 실험에서는 제안 시스템의 구현 가능 여부 뿐 아니라, LP 레코드의 보관 상태 및 마모 정도에 따른 차이 유무를 확인하기 위해 각기 다른 년도에 제작 및 출시 된 다섯 장의 LP 레코드를 선정하였다.

LP 레코드 이미지의 회전을 추가로 활용하는 제안 시스템의 프로토타입 구현을 위해 그림 10에서와 같이 턴테이블의 상단에 카메라를 설치하여 오디오 신호와 함께 영상 정보를 녹화 하였다. 영상을 촬영하는 과정에서는 30fps(frame per second), 854x480 해상도의 웹캠을 사용하였고, 오디오 수음 방식은 앞선 오디오 핑거프린팅 시스템 실험에서와 동일한 설정을 따랐다.

#	Title	Artist	Release
1	Come to Me	France Joli	1979
2	Let's Groove	Earth Wind & Fire	1981
3	Rain Falls	Frankie Knuckles	1991
4	Knee Deep	Maters at Work	2003
5	The Distance	Miguel Migs	2013

표 1. 프로토타입 실험에 사용된 LP 레코드

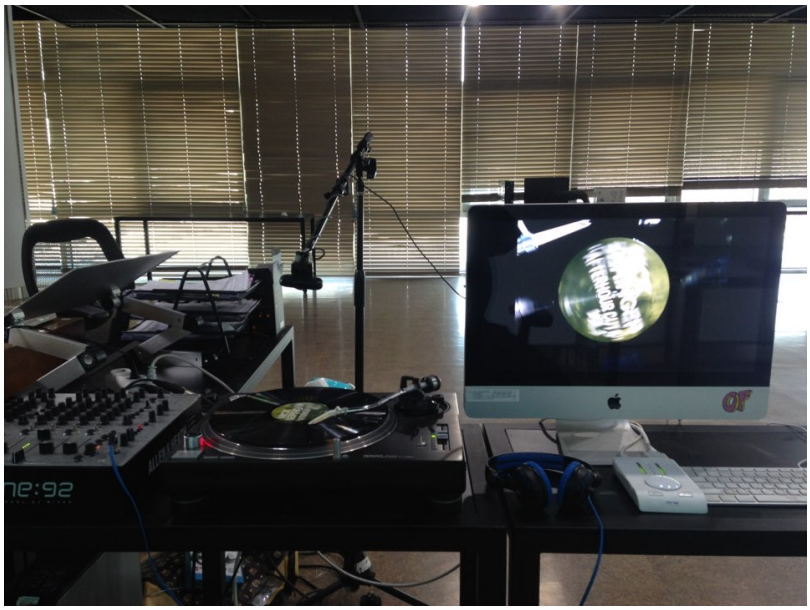

그림 10. 프로토타입 실험 환경

그림 11. 본 실험 환경

3. 1. 3 입력 신호 정렬

제안 시스템을 활용한 본 실험에서는 높은 프레임 레이트를 적용할 수 있는 Apple사의 iPhone6를 사용하여 60fps, 1920x1080 해상도 하에서 실험을 진행하였다 (그림 11). 잡음이 많은 실제 사용 환경을 고려하여 스피커로부터 재생되는 음악을 iPhone6의 내장 마이크를 활용하여 직접 수음 하였다. 본 실험에서 사용된 음원은 제안 시스템의 성능 평가를 위해 유사한 음악적 반복이 많은 다섯 장의 LP 레코드로 구성되었고 (표 2), 이는 프로토타입 실험에서와 동일하게 3초 길이를 가지도록 편집되어 입력 신호로 활용 되었다. 이 때, 정렬 가능한 구간과 불가능한 구간을 자세히 보기 위해 3초의 입력 신호의 시작 지점은 0.1초 간격으로 변화하도록 하였다. 위상의 오차 범위는 ±1° 로 설정했다.

#	Title	Artist	Release
2	The distance	Miguel Migs	2013
3	Knee deep	Maters at Work	2003
5	Sing it back	Deep88	2012
7	Let me	Miguel Migs & Shades of Gray	2013
9	June	Monosoul	2014

표 2. 본 실험에 사용된 LP 레코드

제 2 절 제안 시스템

본 연구의 목표는 메타데이터가 없는 LP 레코드에서도 사용자가 레코드의 정보를 얻을 수 있도록 하는 시스템의 구현이다. 이를 위해 제안하는 시스템의 구조는 크게 두 부분으로 나눌 수 있다. 시스템의 앞단은 입력된 오디오 신호를 이용해 곡을 식별하는 부분이고, 뒷 단은 오디오 신호 및 이미지 회전 정보를 이용해 곡의 실시간 진행 위치를 정확히 찾아내는 부분이다. 입력 오디오 식별의 경우는 기존의 오디오 핑거프린팅 알고리즘으로도 구현이 가능하기에 [4]의 연구를 활용하였다. 시스템의 뒷 부분에 해당하는 실시간 정렬을 위한 제안 시스템은 그림 12과 같은 구조를 가진다.

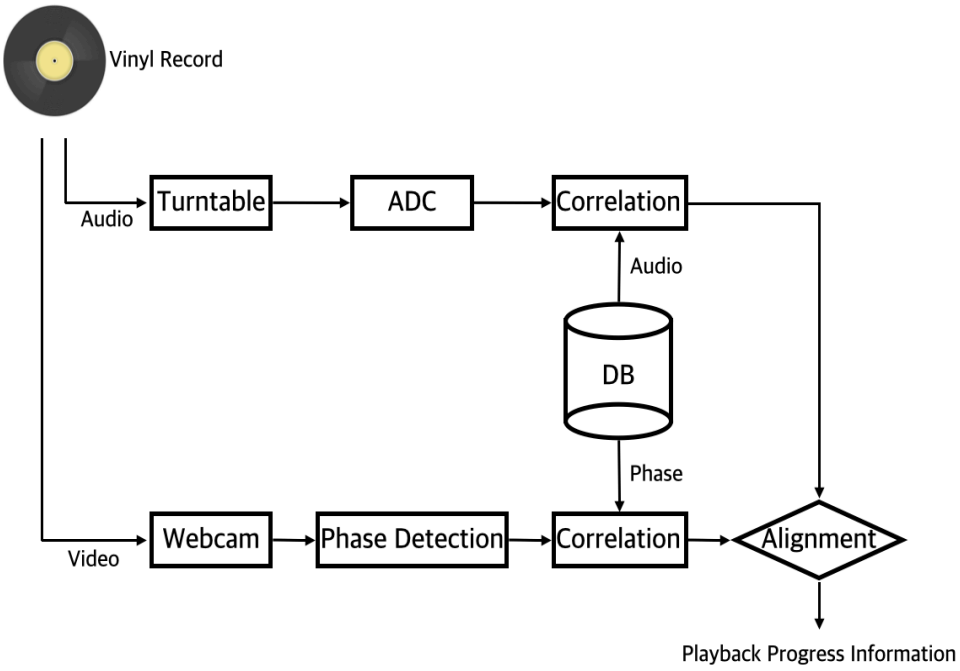

그림 12. 실시간 정렬을 위한 제안 시스템

데이터베이스에는 음원들의 오디오 핑거프린트, 오디오 신호, LP 레코드의 기준 이미지, 그리고 LP 레코드 이미지의 시간 변화에 따른 위상 변화가 저장되어 있다. 위상 변화는 SURF를 활용해 기준 이미지와 각 프레임에서의 이미지 비교를 통해 회전량을 계산한다.

3초의 길이를 가지는 오디오 입력 신호가 들어오면 오디오 핑거프린팅을 통해 식별된 곡 전체 신호를 데이터베이스로부터 참조하여 입력 오디오와의 상호 상관을 계산한다. 이 중에서 상호 상관이 높게 나타난 30개의 후보를 선별하고 각 후보에 해당하는 시간의 위상 정보를 불러온다. 입력 신호의 영상과 데이터베이스의 기준 이미지와의 비교를 통해 입력 신호의 위상을 계산하고, 이를 불러온 30개 후보의 위상 정보와 비교한다. 그 차이가 오차 범위를 벗어나면 후보에서 제외되고, 남은 후보 중에서 상호 상관이 가장 높은 후보를 실시간 진행 위치로 삼고 신호를 정렬한다.

제 4 장 실험 결과

제 1 절 오디오 핑거프린팅

기존 오디오 핑거프린팅 기술의 적용 가능 여부 및 성능 평가를 위해 LP 레코드로 녹음된 50곡의 음원에 대해 오디오 핑거프린트를 추출하여 데이터베이스에 저장한 후, 임의의 구간에서 잘라낸 3초의 오디오 신호를 넣었을 때에 오디오 식별을 정확히 해내는지 실험해 보았다. 기존 오디오 핑거프린팅 논문의 데이터베이스에 비해 그 숫자는 적었지만 성능 확인 결과 모든 곡의 3초 입력에 대해 정확히 식별해 내는 것을 확인할 수 있었다. 따라서 사용자가 가지고 있는 LP 레코드의 오디오 핑거프린트 및 음원 정보 데이터베이스를 사전에 구축해두면 오디오 핑거프린팅 기술을 활용하여 현재 재생 중인 LP 레코드를 식별하고, 곡의 제목, 아티스트, 조성, 빠르기 등 해당 곡의 음원 정보를 다시 사용자에게 제공할 수 있음이 확인 되었다.

제 2 절 입력 신호 정렬

4. 2. 1 SURF 이미지

LP 레코드의 이미지가 입력 신호 정렬에 사용될 수 있는지 여부를 판단하기 위하여 50 곡의 LP 레코드 이미지로부터 SURF 관심점을 추출해 보았다. 대부분의 경우, SURF 관심점을 추출하는 데에 문제가 없었기에 회전 정보를 추출할 수 있었지만 SURF 관심점을 추출할 수 없는 경우들이 발견되었다.

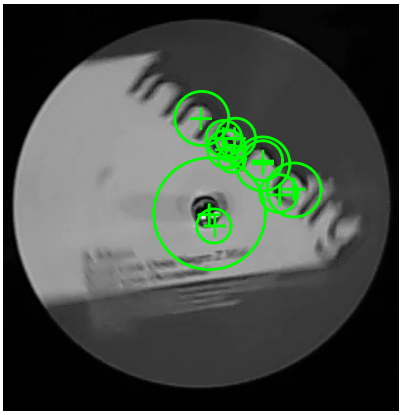

그림 13. 충분한 SURF 관심점

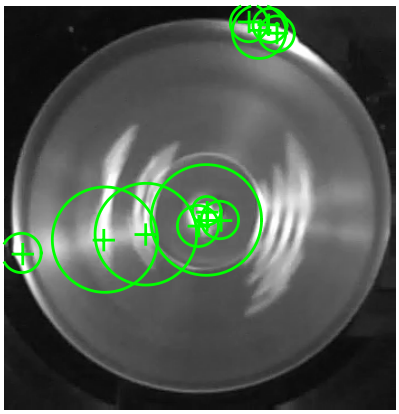

그림 14. 불충분한 SURF 관심점

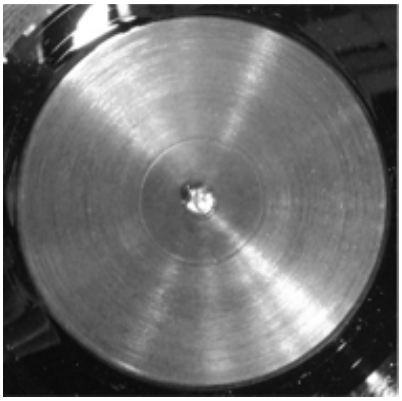

그림 15. 제안 시스템에 활용될 수 없는 예

그림 13은 충분한 양의 SURF 관심점이 추출되어 이미지 위상 계산에 활용될 수 있는 예 이다. 하지만 그림 14의 경우는 충분하지 못한 조도로 인해 이미지의 경계점이 불분명해져 SURF 관심점을 충분히 찾지 못한 경우이다. 이러한 경우, 충분한 조도 하에서 다시 영상을 촬영하였을 때 그 문제가 해결됨을 확인할 수 있었다. 웹캠의 거리가 멀어 이미지를 정확히 잡지 못한 경우에도 비슷한 문제가 발생하였으나 웹캠의 거리를 좁히고 충분한 조도를 확보함으로써 문제를 해결할 수 있었다. 그림 15는 LP 레코드 이미지가 단색으로 되어 SURF 관심점을 찾을 수 없는 경우이다. 이러한 경우는 제안하는 시스템에서 활용될 수 없다. 또한 실험에 사용된 50곡에 포함되지는 않았지만 상하좌우가 대칭인 이미지의 경우 역시 회전 정보를 활용하기에는 부적합 할 수 있을 것으로 예상된다.

4.2.2 프로토타입 실험

앞선 실험에서 발견된 문제를 방지하기 위해 충분한 조도와 웹캠 거리를 확보한 상태에서 표 1의 LP 레코드를 각각 2회씩 녹화 및 녹음했다. 다섯 장의 레코드는 각기 다른 시기에 제작되었고 다른 장르적 특성을 지니고 있다. 1회분의 녹화 및 녹음 정보는 데이터베이스로 사용되었고 다른 1회분의 녹화 및 녹음 정보는 3초의 길이를 가지도록 분할하여 입력 신호로 사용하였다. 데이터베이스로 사용된 영상의 첫 프레임은 변화하는 위상 계산을 위한 기준 이미지로 사용되었다.

위상을 계산하는 첫 실험을 진행한 결과, 기준 이미지와의 각도 차이를 구하는 과정에서 오차가 발생하여 그 변화가 선형적이지 않음이 발견되었다. 몇몇 SURF 관심점이 잘못 잡히거나 잡지 못하는 경우에 계산된 위상에 오차가 존재하였다. 그림 16은 측정된 위상이 안정적으로 일정하게 변하는 모습을 보이고 있다. 30fps에서의 3초간의 변화이므로 90 프레임 동안 0° 에서 360° 까지 일정하게 증가하는 패턴을 반복한다. 하지만 그림 17에서는 위상을 잘못 찾는 프레임이 일부 존재함을 알 수 있다. 두 번째 실험에서는 이러한 오차를 줄이기 위해 인접한 프레임의 각도를 참조하여 이를 보정하였다.

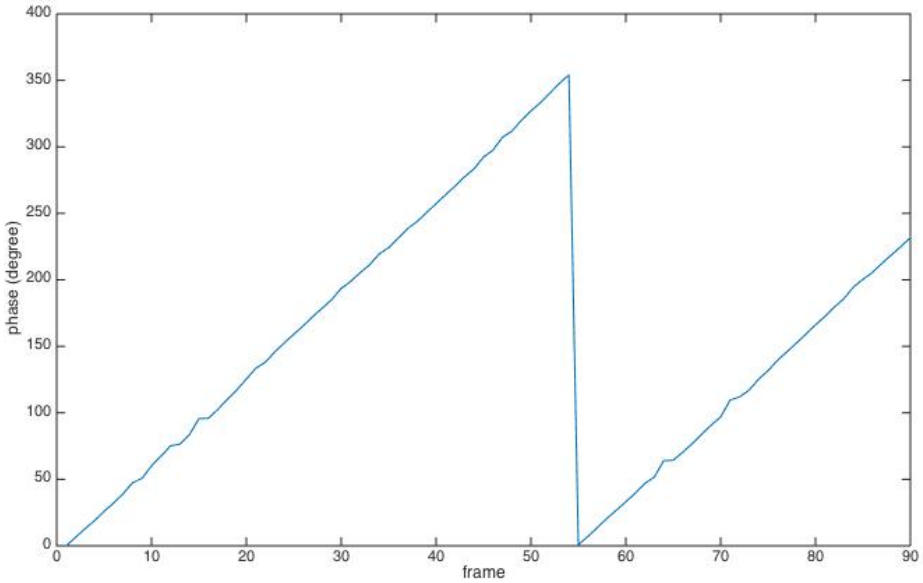

그림 16. 안정적으로 변화하는 LP 레코드 위상

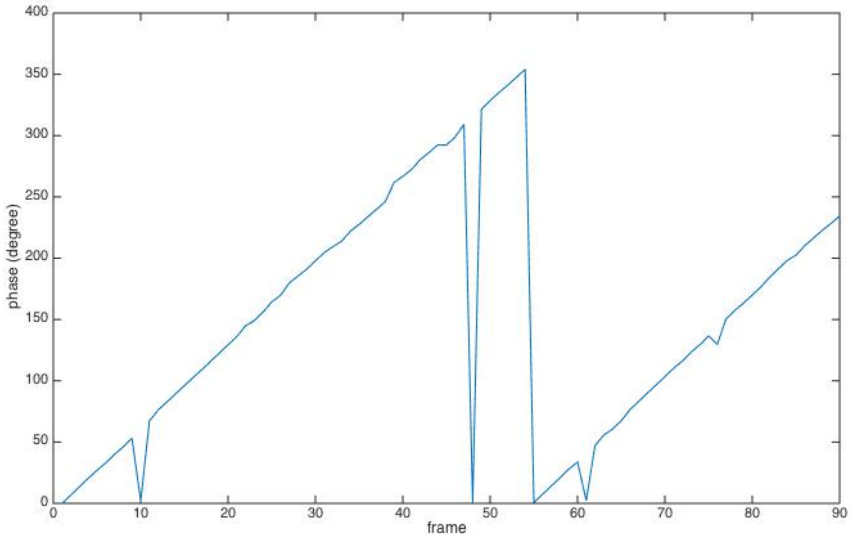

그림 17. 오차가 존재하는 LP 레코드 위상

두 번째 실험에서는 입력 신호와 데이터베이스의 오디오 상호 상관이 높은 30개의 후보를 추출하고 각 위치에서의 위상 차이를 비교하여 위상 차이가 4° 이상 나는 경우, 실시간 정렬 후보에서 제외 하였다. 30 fps의 웹캠을 활용하여 33 1/3rpm의 속도로 회전하는 이미지를 촬영하는 경우 각 프레임에서는 약 6.67° 위상 변화가 일어난다. 하지만 이 때, 오디오 신호는 영상 신호보다 정밀한 6000Hz의 샘플링 레이트를 가지기에, 실시간 정렬 후보의 시간 정보를 영상에 적용하는 경우, 가까운 프레임의 위상을 가져옴으로써 6.67° 의 오차가 추가로 발생하는 현상이 발견되었다.

프레임 레이트로 인해 생기는 오차 및 위상 계산에서의 오차를 해결하기 위해 SURF를 통해 찾은 정보를 보간법(interpolation)을 통해 오디오 신호와 동일한 샘플링 주파수인 6000Hz의 정보로 변환하여 실험

을 진행 하였다. 오디오 신호의 상호 상관만을 이용해 입력 신호를 정렬한 경우, 다섯 장의 LP 레코드에서 102개의 잘못된 정렬이 있었다. 그러나 위상 정보를 활용한 경우에는 이 중 76개를 맞게 정렬할 수 있었다. 그림 18의 상단은 오디오의 상호 상관을 통해 추출한 30개의 후보를 나타낸다. 이들은 비슷한 값의 상호 상관을 가지고 있다. 그림 18의 두 번째 그림은 오디오 상호 상관이 가장 높은 한 지점을 나타낸다. 그림 18의 하단 그림은 30개의 후보 중 위상 정보가 오차 범위 안에 있고 상호 상관이 높은 정렬 지점을 나타낸다. 이렇듯 오디오 상호 상관에서의 잘못된 정렬이 위상 변화를 참조함으로써 보정되는 것을 볼 수 있다.

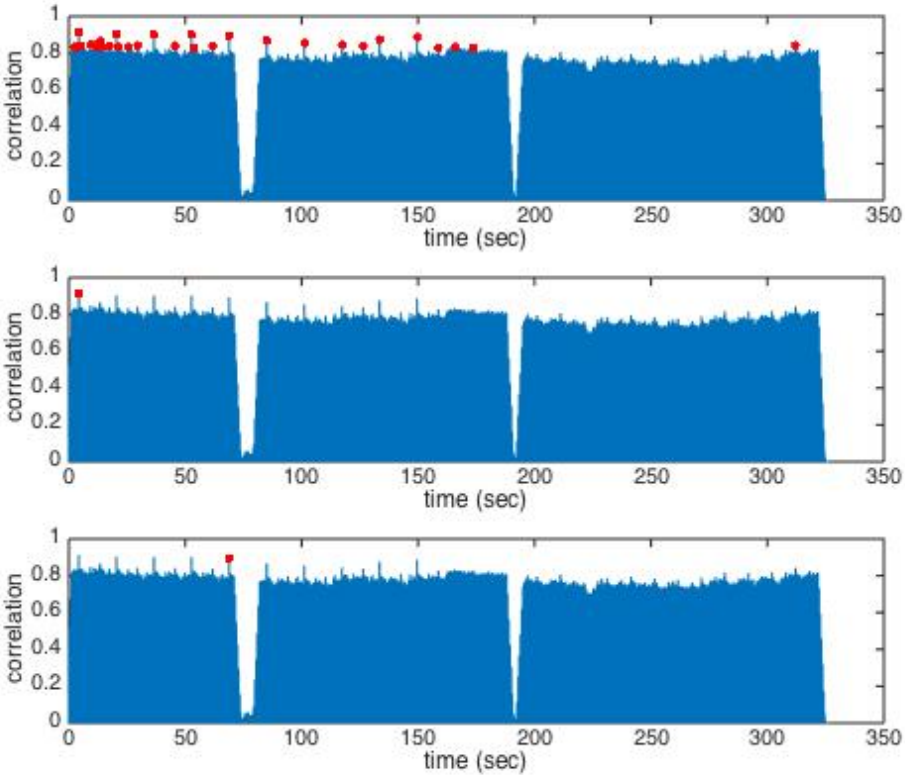

그림 18. 70초-73초 구간의 입력 신호를 넣었을 때의 오디오 상호 상관

실험	정확도 (%)
오디오 상호상관	94.49
프로토타입(30fps)	95.35

표 3. 오디오 상호 상관을 사용했을 때의 정렬 정확도와 프로토타입 실험에서의 정렬 정확도 비교

위상 정보를 추가로 사용했음에도 오류의 25.5%는 보정에 실패했다. 또한 오디오 상호 상관 만을 활용했을 때 맞게 정렬됨에도 불구하고 위상 오차로 인해 후보에서 누락되는 경우도 존재하였다. 이는 위상을 계산하는 과정에서 오차가 발생 하였거나, 보간법으로 구한 값이 오차를 보정하지 못한 경우, 혹은 정답에 해당하는 후보가 아님에도 불구하고 위상이 오차 범위 안에 들어 정렬되는 경우 등 이었다. 이로인해 표 3에서 볼 수 있듯 뚜렷한 성능 개선이 이루어지지 않음을 확인할 수 있었다.

4. 2. 3 디지털 음원과의 비교

실험에서 사용된 LP 레코드는 2회에 걸쳐 녹음 및 녹화 되었지만 동일한 레코드를 사용하였다. LP 레코드는 물리적으로 회전을 하면서 그 정보를 읽어들이기 때문에 동일한 레코드에서는 동일한 잡음이나 오디오 특성이 나타나게 된다. 이러한 잡음이나 특성으로 인해 오디오

상호 상관을 계산하였을 때, 동일한 반복이 있는 다른 구간들에 비해 해당 위치에서만 상호 상관이 높게 나타나는 경우가 있었다. 실제 사용자가 시스템을 사용하는 경우, 데이터베이스는 같은 오디오 신호를 담고 있는 다른 레코드로 구축되어 있거나 디지털 음원을 통해 구축되어 있을 가능성이 높다. 이 경우의 오차를 알아보기 위해 앞선 프로토타입 실험의 오디오 신호를 디지털 음원으로 교체하여 동일한 실험을 진행해 보았다.

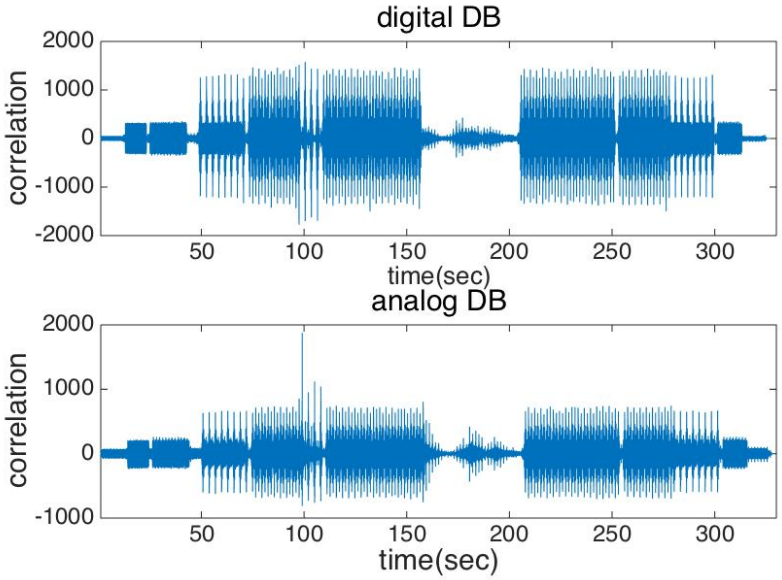

그림 19. 디지털 음원과 비교했을 때의 상호 상관과 아날로그 음원과 비교했을 때의 상호 상관

그림 19의 상단은 LP 레코드 입력 신호와 디지털 음원과의 상호 상관을 나타낸다. 비슷한 패턴의 반복이 많은 곡이기에 비슷한 값의 상호 상관을 가지는 후보들이 많이 존재한다. 하지만 비교 대상을 기존의 LP 레코드 음원으로 했을 경우에는 그림 19의 하단과 같이 하나의 구간에서 상호상관이 높게 나타난다. 이러한 현상은 LP 레코드가 오래되

어 흠집이 많거나 마모된 경우에 더욱 자주 나타났다. LP 레코드에 있는 흠집이 고유의 오디오 특성으로 작용하기 때문에 동일한 신호를 가졌던 다른 구간과 구분되는 신호를 띄게 된 것이다.

오디오 상호 상관만을 이용했을 때 잘못된 정렬이 6개에 불과했던 5번 레코드는 디지털 음원을 비교 대상으로 삼은 경우, 53개의 잘못된 정렬을 보였다. 프로토타입 시스템은 그 중 33개를 맞는 위치로 다시 정렬 하였다.

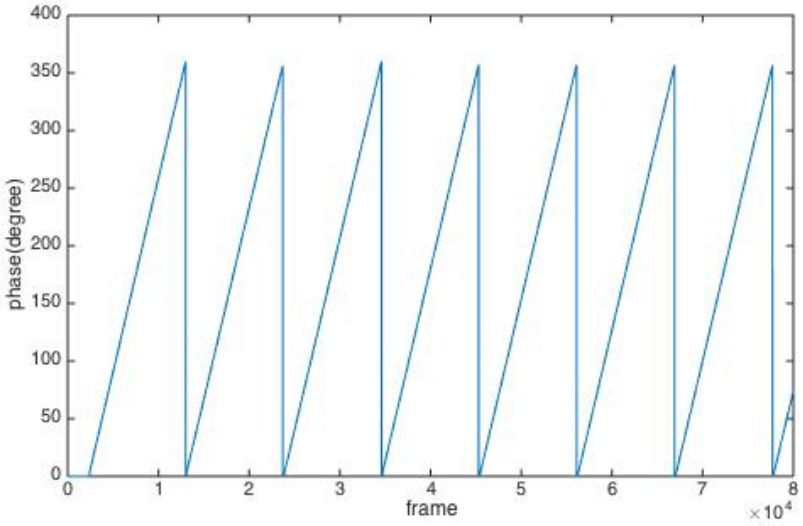

그림 20. 안정적으로 변화하는 위상 (60 fps)

4. 2. 4 입력 신호 정렬 (60 fps)

본 실험에서는 LP 레코드 이미지의 위상 정보 활용이 유용함을 보이기 위해 마모가 덜한 최신의 LP 레코드를 보다 높은 해상도, 높은 프레임 레이트 하에서 촬영하였다. 표 2의 다섯 장의 LP 레코드를 iPhone 6를 활용하여 60fps로 촬영하였고, SURF를 활용하여 위상을 측

정한 결과, 그림 20에서와 같이 안정적으로 위상이 변하는 것을 확인할 수 있었다. 프로토타입 실험에서의 보간법을 수행하지 않았음에도 대부분의 구간에서 위상은 안정적으로 측정되었다.

실험 결과, 17009개의 입력 신호 중에서 오디오 상호 상관만을 사용한 경우 3098개의 잘못된 정렬이 일어났다. 이는 표 4에서와 같이 81.79%의 정확도를 나타낸다. 여기에 제안한 시스템을 적용하는 경우, 잘못된 정렬 중 1821개를 맞는 위치로 정렬하여 92.49%로 정확도가 증가함을 확인할 수 있었다. 위상 정보를 활용함에도 잘못된 정렬이 일어나는 경우는 크게 세 가지 경우 였다. 하나는 입력 신호와 유사한 신호가 곡 안에 다수 존재함으로 인해 입력 신호와 상호 상관이 높은 30개의 정렬 후보에 실제 입력 신호의 재생 구간이 들어가지 못하는 경우였다. 이러한 경우, 후보에 입력 신호가 없기 때문에 위상 정보를 참조할 수 없어 제안하는 시스템을 통해서도 보완이 불가능 하다. 잘못된 정렬의 또다른 경우는 반복되는 구간의 위상이 다른 위치에서도 동일하게 나타나는 경우이다. 오차 범위가 ±1° 이기에 그 확률은 낮지만 그림 21, 22과 같이 다른 구간에서 위상이 같은 경우, 제안하는 시스템을 활용하여도 맞는 정렬을 하지 못할 수 있다. 실제로 5번 레코드에서 이러한 현상이 있어 특정 구간에서 정확도가 낮음을 볼 수 있었다.

실험	정확도 (%)
오디오 상호상관	81.79
제안 시스템(60fps)	92.49

표 4. 오디오 상호 상관을 사용했을 때의 정렬 정확도와 제안 시스템의 정렬 정확도 비교

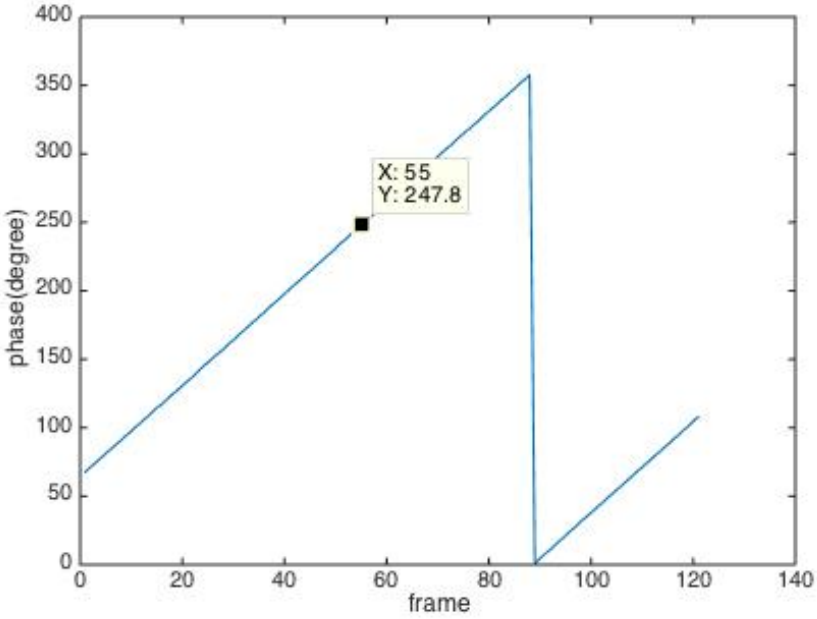

그림 21. 오디오 상호 상관이 높은 55초 구간에서의 위상

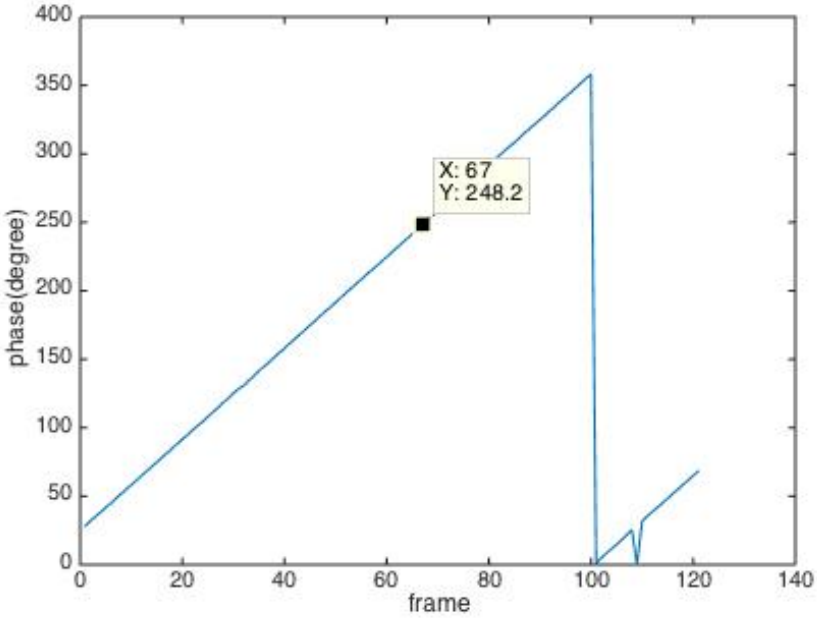

그림 22. 오디오 상호 상관이 높은 67초 구간에서의 위상

반복 구간에서 위상이 겹치는 5번 레코드를 제외한 4개의 레코드의 경우, 오디오 신호만을 활용한 정렬의 정확도는 84.84% 였고, 제안한 시스템을 활용한 정렬의 정확도는 96.76% 였다. 이렇듯 오디오 신호와 위상 정보가 동일함으로 인해 발생하는 오류를 개선하기 위해서는 LP 레코드의 또다른 특징을 활용할 수 있을 것으로 예상된다. 턴테이블을 활용하여 LP 레코드를 재생하는 경우, 카트리지의 위치가 곡 진행에 따라 점차 중앙을 향해 이동하게 된다. 이 카트리지의 위치를 추가적으로 활용하게 되는 경우, 오디오 신호와 위상이 동일한 경우를 구분해낼 수 있다. 제안 시스템이 정확한 정렬에 실패한 세 번째 경우는 SURF를 활용한 위상 추정에서의 오차로 인한 경우이다. 그림 23에서와 같이 SURF 관심점의 잘못된 연결로 인해 위상 추정에 오차가 발생하는 경우, 잘못된 위치에 신호가 정렬됨을 볼 수 있었다.

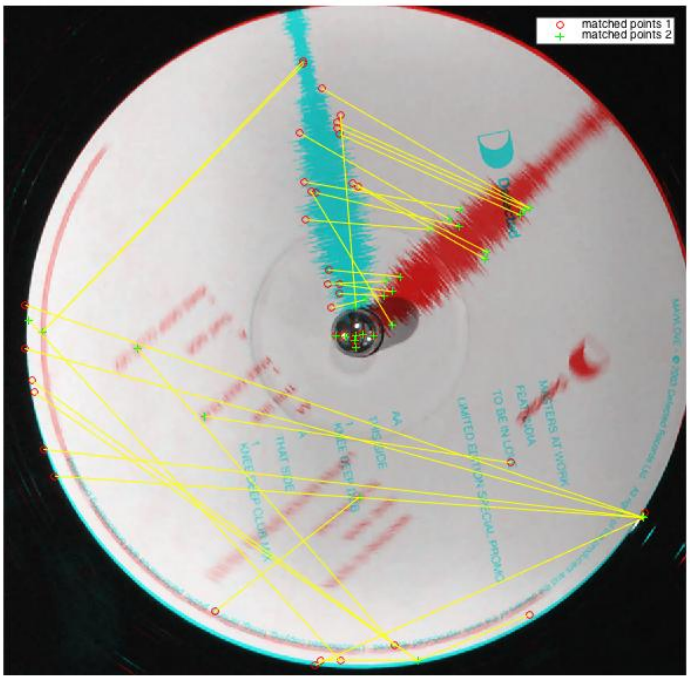

그림 23. SURF 관심점이 잘못 연결된 예

제 5 장 결론

본 논문은 오디오 핑거프린팅 기술과 영상 처리 기술인 SURF를 활용하여 메타데이터가 없는 아날로그 매체인 LP 레코드에서도 사용자에게 실시간 음원 정보를 제공할 수 있는 시스템을 제안하였다. 입력 신호를 통해 음원 정보가 담긴 데이터베이스에 접근할 수 있도록 오디오 핑거프린팅을 활용한 간단한 실험을 진행하였고, 3초의 오디오 신호 입력 만으로도 음원을 정확히 식별하여 사용자에게 곡 제목, 아티스트, 앨범 명 등의 음원 정보를 제공할 수 있음을 확인 하였다.

그러나 오디오 신호 입력 만으로는 실시간 재생 구간을 판단하는데에 오류가 발생하기 때문에, 이를 극복하기 위해 제안하는 시스템에서는 턴테이블의 상단에 카메라를 설치해 LP 레코드가 회전하는 영상 정보를 함께 활용하도록 하였다. 기존 오디오 상호 상관만을 사용했을 때, 81.79%의 정확도를 보이던 신호 정렬은 제안하는 시스템을 통해 92.49%의 정확도를 보이도록 개선되었다. 제안된 시스템을 활용하면 사용자에게 실시간 음원 정보를 제공할 수 있을 뿐 아니라, DJ 들의 공연에 유용하게 사용될 수 있을 것으로 기대된다.

실험에서 사용된 오디오 및 영상 신호는 실험자가 직접 녹음 및 녹화 하기에 실시간 동작 시스템이 아니다. 시스템에 대한 정확한 평가를 위해선 실시간 시스템으로 구현된 뒤에 DJ 및 일반인에 대한 사용자 평가가 필요하다. 여기서 예상되는 문제점은 카메라 라는 별도의 장치를 추가해야 한다는 번거로움과 손으로 사용자가 LP 레코드의 이미지를 가리는 경우가 발생할 수 있다는 점이다. 카메라를 턴테이블의 상단에 위치 시키지 않고 하단에 위치 시켜 뒷면의 이미지를 촬영하는 경

우, 외부의 조명에 영향을 받지 않는 안정적인 조도를 확보할 수 있고, 사용자가 보다 자유롭게 시스템을 활용할 수 있을 것으로 기대된다. 또한 실시간 구현의 경우 연산량을 고려하여야 하기 때문에 이미지의 해상도, 오디오 입력 신호의 샘플링 주파수에 따른 시스템의 성능 변화도 확인이 필요하다.

연구에서 실시된 실험은 회전하는 LP 레코드의 위상을 측정하기 위해 영상 처리 기술 SURF를 사용하였다. SURF는 조명, 카메라 각도, 이미지의 크기, 회전 등 외부 요인에 의한 왜곡에 대해 강인하다는 특징을 가지지만 이는 추가적인 연산량이 발생함을 의미한다. 조도, 카메라의 각도, 카메라의 거리 등을 고정시킨 상태에서 실험을 한다면 보다 단순화된 영상 처리 알고리즘 만으로도 빠르게 위상 추적을 할 수 있을 것으로 기대된다.

본 연구는 음악 소비자의 LP 레코드 이용에서 발생하는 불편함을 오디오 및 이미지 신호 처리 기술을 활용해 극복하려한 연구로서, LP 레코드와 메타데이터를 연결하고, DJ 공연에 활용 가능한 기술적 기반을 제공함에 그 의의를 가진다.

참고문헌

[1] Nielsen SoundScan. 〈http://www.nielsen.com〉

[2] Greg Sandoval. “Amazon vinyl sales up 745 percent since 2008, but it won’t save music.” 〈http://www.theverge.com〉

[3] D. Bartmanski, and I. Woodward: Vinyl: The Analogue Record in the Digital Age, Bloomsbury Publishing, 2015.

[4] J. Haitsma, and T. Kalker. "A highly robust audio fingerprinting system with an efficient search strategy." *Journal of New Music Research* 32.2 (2003): 211-221.

[5] A. Wang: "An Industrial Strength Audio Search Algorithm," *ISMIR*, 2003.

[6] Cano, Pedro, et al: "A review of audio fingerprinting." *Journal of VLSI signal processing systems for signal, image and video technology* 41.3 (2005): 271-284.

[7] A. Camarena-Ibarrola, and E. Chávez. "A robust entropy-based audio-fingerprint." *Multimedia and Expo, 2006 IEEE International Conference on*. IEEE, 2006.

[8] J. Seok, J. Hong, and J. Kim. "A novel audio watermarking algorithm for copyright protection of digital audio." *etri Journal* 24.3 (2002): 181-189.

[9] D. Milano: "Content control: Digital watermarking and fingerprinting."*Rhozet, Harmonic, White Paper. http://www. rhozet. com/whitepapers/Fingerprinting Watermarking. pdf* (2012).

[10] RIAA-IFPI. Request for information on audio fingerprinting technologies, 2001.

〈http://www.ifpi.org/site-content/press/20010615.html〉

[11] A. Camarena-Ibarrola, and E. Chavez: “Real Time Tracking of

Musical Performances," Advances in Soft Computing, pp. 138-148, 2010.

[12] S. Dixon, and G. Widmer: "MATCH: A music Alignment Tool Chest," Proceedings of the International Symposium on Music Information Retrieval, pp. 492-497, 2005.

[13] S. Ewert, M. Muller, and P. Grosche: "High Resolution Audio Synchronization Using Chroma Onset Features," In Acoustics, Speech and Signal Processing, pp. 1869-1872, 2009.

[14] N. Hu, R. B. Dannenberg, and G. Tzanetakis: "Polyphonic Audio Matching and Alignment for Music Retrieval," IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2003.

[15] O. Izmirli, and R. B. Dannenberg: "Understanding Features and Distance Functions for Music Sequence Alignment," Proceedings of the International Symposium on Music Information Retrieval, pp. 411-416, 2010.

[16] M. Muller: "Dynamic Time Warping," Information Retrieval for Music and Motion, pp. 69-84, 2007.

[17] H. Bay, T. Tuytelaars, and L. Van Gool: "SURF: Speeded Up Robust Features," In Computer vision - ECCV, pp. 404-417, 2006.

Abstract

Vinyl Record Information Retrieval via Audio Fingerprinting and Image Processing

Minz Sanghee Won

Department of Transdisciplinary Studies

The Graduate School

Seoul National University

This paper presents a vinyl record information retrieval system using Speeded Up Robust Features (SURF) for ordinary music listeners and disc jockeys. When users listen to music with a digital music player, the device provides real time information (e.g., the album title, the name of the artist, length of the song, and playback progress) with the music. This information is considerably helpful for disc jockeys as well as ordinary music listeners. However, a turntable, which is a device for playing vinyl records, cannot provide any real-time information for users. Through conventional audio fingerprint algorithms, it is possible

to access to a vinyl record's metadata in a pre-constructed database. For more instantaneous information such as the playback progress, existing audio alignment algorithms can be helpful. However it is highly likely that several competing locations will be found, particularly for popular music because it consists of many repetitions that are nearly the same. In this paper, I propose a precise audio alignment system using a unique characteristic of vinyl records – its physical rotation. Given an audio input as a query, my proposed system first identifies the song by audio fingerprint system. And the system finds the candidate locations using the conventional audio alignment technique. Then pinpoints the precise location of the query by referring to the current phase of the rotating image. Through the experiments I demonstrate that the proposed system can successfully find the exact location of the query audio.

Keywords : Music Technology, Vinyl Record, Music Alignment, Music Information Retrieval, Disc Jockey

Student Number : 2013-22409