

## ListenPhoto: 자동 슬라이드 쇼 배경음악 추천 시스템

서중훈<sup>○</sup> 김동철 한탁돈

연세대학교 컴퓨터과학과 미디어시스템연구소

{jonghoon.seo<sup>○</sup>, dckim, tackdon}@msl.yonsei.ac.kr

### ListenPhoto: Automatic Music Annotated Photo Browsing System

Jonghoon Seo<sup>○</sup> Dongchul Kim Tackdon Han  
Yonsei University

#### 요 약

디지털 카메라의 보급과 U.C.C. 개념의 등장으로 인하여 개인이 생산하는 사진의 양도 많아지고, 이를 감상하는 시간도 많아졌다. 사진이란 추억을 회상하는데 하나의 포인트 역할을 하는 매체이다. 따라서 인간의 감성적인 부분을 많이 자극되는데, 이 과정에서 추가적인 자극을 줌으로써 사용자의 감성을 더욱 자극할 수 있다. 하지만, 기존의 연구들은 디지털 사진을 검색하고 관리하는데 초점이 맞추어져 있었다. 본 논문에서는 사용자가 사진 앨범을 감상할 때 개별 사진을 분석하여 전체 사진과 가장 잘 어울리는 배경음악을 추천/재생하는 ListenPhoto 시스템을 제안한다. 그리고, 시스템의 필요성에 대한 평가와 시스템 성능에 대한 평가 두 부분으로 나누어 평가를 진행하였다. 사용자들은 이 시스템을 통하여 풍부하게 디지털 사진 생활을 즐길 수 있게 될 것이다.

#### 1. 서 론

디지털 가전기기들의 증가로 인하여 개인이 생산하는 정보의 양이 급격하게 증가하고 있다. 특히 디지털 카메라라는 가격 인하와 성능 향상 등으로 인하여 널리 보급되었고, 이에 따라 개인 사용자가 생산하는 디지털 사진의 양이 급속도로 증가하였다. 개인이 생산한 디지털 사진의 증가로 데이터의 저장과 관리, 감상이 중요한 문제로 부각되었고 이를 효율적으로 관리하고 검색하기 위한 편리한 인터페이스 및 검색 시스템들이 연구되어 왔다.

그러나 기존의 시스템들은 데이터를 관리하고 검색하는데에 주로 초점이 맞추어져 있고 감상에 대한 연구는 미비한 상태였다. 사진이라는 매체의 중요한 목적은 기록이고 이것을 보았을 때 찍은 당시의 감성의 회상이기 때문에 감성적 자극제로써의 역할이 중요하다. 이 때 여러 감각이 복합적으로 사용자에게 감성적 자극을 미친다. 따라서 사진의 감상에 있어서도 시각 이외의 여러 감각을 적절하게 자극함으로써 사용자에게 최고의 감성적인 자극을 전달해 줄 수 있다.

본 논문에서는 사진을 보는데 있어서 영상 분석을 통하여 적합한 배경음악을 재생해 줌으로써 시각과 청각을 자극 하도록 하였다. 사용자가 재생할 사진 앨범을 선택하면 시스템은 각 사진들의 영상 분석과 여러 상황정보를 파악하고 학습된 데이터에 의해 가장 적합한 음악을

재생함으로써, 사용자에게 적절한 다감각적 자극을 제공한다. 이로써 사용자는 보다 풍부하게 감성적 자극을 받는 사진 감상을 할 수 있게 된다.

2장에서는 현재 관련 산업과 학계의 동향을 조사하였고 3장에서는 제안하는 ListenPhoto 시스템에 대해 기술하였으며, 4장에서는 사용자 평가를 하고 5장에서 시스템의 활용 분야를 제안하면서 결론을 맺는다.

#### 2. 관련연구

대부분의 기존 연구는 방대한 양의 사진의 효율적인 관리에 중점을 두고 있다. 즉, 사진의 그룹핑이나 태깅, 또는 내용 기반 검색 등의 영역이 주로 연구되고 있다. 그에 반해 직접 사진을 이용하게 되는 경우인, 감상 단계에 대한 연구는 거의 이루어지지 않고 미흡했던 실정이다.

상업적인 영역에서는 사용자들의 반응이 프로그램의 판매에 직접적으로 연관되므로 이러한 감상 단계에 대한 고민의 흔적이 많이 보였다. 특히 적합한 배경음악의 조합이 사용자들의 사진 앨범 감상에 정서적으로 도움을 준다는 필요성을 느끼고 많은 시스템에서 배경음악 삽입 기능을 제공하고 있다.

하지만, Preclick Gold Photo Organizer의 경우, 아직 배경음악 삽입 기능조차 제공되지 않았고, Picasa나 Motion Browser, Digital Photo Slide Show, FastStone

Image Viewer 등은 사용자가 폴더를 지정하면, 시스템은 이 폴더에서 임의의 순서대로 (가령 알파벳 순서) 배경음악을 선택하여 사진 슬라이드 쇼 감상 시 첨가한다. 이러한 방식의 경우, 한번만 자신의 음악 파일 폴더를 지정하면 되기 때문에 사용자의 노력이 거의 없는 반면, 시스템 임의로 배경음악을 폴더에서 선택하기 때문에 보통의 경우 감상하고자 하는 사진 앨범과 동떨어진 배경음악이 삽입되게 되고, 이는 사용자의 사진 감상에서 감동을 떨어뜨리게 된다.

또 다른 방법으로, ACDSee와 마이크로소프트사의 Photo Story 3와 같은 경우에는, 사용자가 사진 슬라이드 쇼마다 수동으로 배경음악을 선택하는 방식을 선택하였다. 이러한 시스템은, 사용자들에게 슬라이드 쇼에 가장 적합한 배경음악을 선택하도록 함으로써 앞선 방식들이 가지는 슬라이드 쇼와 배경음악의 비매칭 문제를 해결하였지만, 이 방식은 사용자들이 슬라이드 쇼마다 배경음악을 입력하여야 하므로 사용자 노력이 크게 증대된다는 단점을 가지고 있다.

이렇듯, 사진 감상에서의 배경음악의 중요성을 인지하고 이에 대한 지원이 이루어지고 있지만, 주로 슬라이드 쇼와 배경음악 간의 비매칭 문제가 있거나, 사용자의 노력이 커지는 문제점이 있었다. 본 논문에서는 이러한 문제들을 해결하기 위하여 두 방법의 장점을 조합한다. 사용자는 단 한번의 음악 폴더 지정으로 사용자 노력을 최소화하고, 이를 시스템이 분석하여 슬라이드 쇼에 가장 적합한 음악을 추천함으로써, 감상하려는 슬라이드 쇼와 배경음악간의 불일치 문제를 해결하였다.

상업적인 흐름에 반해, 전체적으로 학술적으로는 사진 감상, 특히 배경음악과 사진 감상에 관한 연구는 거의 진행되지 않고 있다. [4]는 음악의 비트를 분석하여 비트가 변화하는 지점에서 슬라이드 쇼의 사진을

전환함으로써 전체적인 슬라이드 쇼와 배경음악이 유기적으로 전환되는 효과를 가져왔으나, 음악의 선곡은 여전히 사용자의 몫이었다. 또한, [2], [3]은 [4]와 같은 방법으로 배경음악을 제공하나, 이 슬라이드 쇼를 타일 방식으로 보여줌으로써 새로운 감상 방법을 제시한 것이 특징이다. [5]의 경우, 높은 연관관계를 가지는 음악과 사진을 연결하여 보여준다는 점에서 제안하는 시스템과 비슷한 점이 있으나, 텍스트 기반인 가사를 기반으로 하여 음악을 선택하였다는 차이점이 있다.

### 3. 제안하는 시스템

ListenPhoto 시스템은, 크게 3 가지 단계로 구분된다. 그림 1 에서는 이에 대한 전체적인 흐름을 보여준다.

- 1) **Machine Learning Stage:** 실제 사용을 위한 사전 단계로, 시스템은 사용자에게 서비스를 제공하기 전에 시스템 학습 단계를 거친다. 이는 미리 선정된 대표적인 카테고리에 해당되는 학습 이미지들을 이용하여 구분자를 만드는 단계이다. 이 단계에서는, 웹 2.0의 개념을 적용하여 사용자들에 의해서 광범위하게 분류되어 있는 웹의 사진들을 이용한다.
- 2) **Recommendation Stage:** 사용자가 감상할 사진 앨범을 선택하여 '자동 음악 선곡'을 실행하면 시스템은 각 사진의 특징점을 추출하고, 특징점과 연관되는 태그를 찾는다. 찾아진 태그들은 찾아진 개수 순서로 연관된 음악을 Tag Matcher에서 찾은 후 해당되는 배경음악을 추천하게 된다.
- 3) **User Feedback Stage:** 사용자는 추천된 음악 리스트에서 최상위 음악이 아닌 다른 음악을 선택할 수 있다. 이는 시스템의 태그 매칭이 잘못 되었음을 말하며, 이는 가중치로 Tag Matcher에 반영되어 사용자

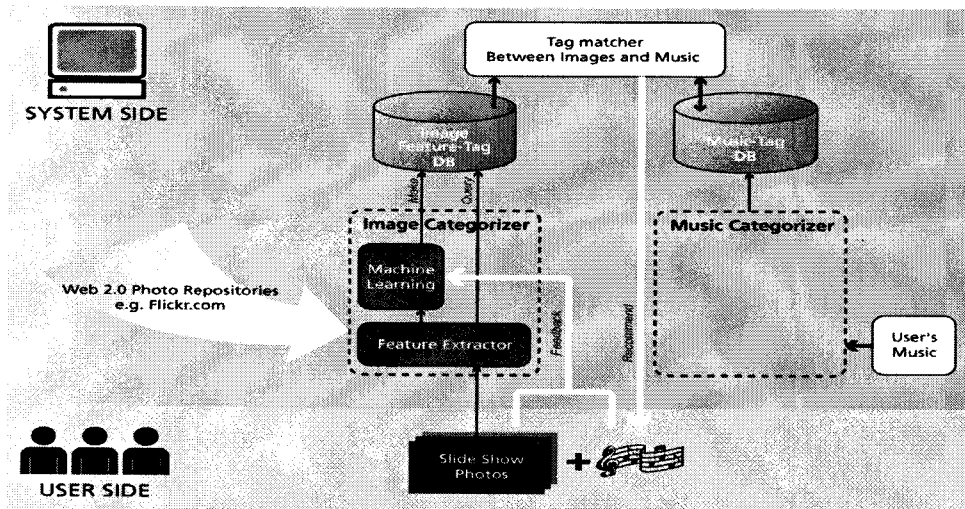


그림 1. ListenPhoto 시스템 흐름도

의 특성을 반영할 수 있도록 한다.

### 3.1. Machine Learning Stage

이 단계는, 미리 선택된 대표적인 카테고리에 해당되는 테스트 이미지들을 통하여 구분자를 정의하는 단계이다. 우선, 본 논문에서는 제안하는 컨셉의 증명을 위하여 3개의 대표 카테고리만을 사용하였다. - 결혼식, 여름 휴가, 졸업 - 이는 매칭될 음악의 분류의 용이성에 근거하여 선정되었다.

하지만, 확장성을 위하여 사용자가 자신의 카테고리를 등록하고 이를 학습시킬 수 있는 기능을 보완하였다. 사용자가 새로운 카테고리를 입력하면, 시스템은 웹2.0의 OpenAPI 기술을 이용하여 이에 해당되는 이미지들을 실시간으로 사용자에게 제시하고, 이 중, 카테고리를 잘 표현하는 20장의 학습 이미지를 선택함으로써 새로운 카테고리를 학습할 수 있다.

이렇게 선택된 학습 이미지들은 먼저, 특징점을 추출하게 되고, 이 특징점과 입력된 카테고리를 이용하여 학습을 통하여 구분자를 만들게 되고 이는 '사진 특징-태그 DB'에 저장된다. 또한, 본 논문에서는 음악 구분은 수동으로 이루어지도록 했다. 이는, 음악의 경우 사진에 비하여 변동이 적기 때문에 우선 수동 분류 방식을 적용하였다.

#### 3.1.1. Feature Extraction Algorithm

사진에서 특징점을 이루는 정보는 다양하다. 이 정보들은 사진의 문맥(Context) 정보라 할 수 있다. 이 시스템에서 사용 가능한 문맥 정보는 아래와 같다.

표 1. 사진에 연관된 문맥 정보들

정보의 출처	사진 문맥 정보
영상 분석	색상 정보, 특징점 분포, 분위기, 등
EXIF	계절, 시각 등
연관 텍스트	사용자에 의한 태그 정보, 저장된 폴더명, 사진 관련 글 등
GPS	장소, 위치 등

본 논문에서는 간단한 구현을 위하여 사진 분석을 통한 특징점만을 이용하였다. 추후의 구현에서는 관련 모든 상황 정보를 이용하는 것이 목표이다.

사진의 특징을 추출하는 방법에 있어서는 속도 향상을 위해 내용기반 자동 사진 방향 판별 알고리즘[1]을 단순화하여 접근하였다. 원본 이미지에서 간단한 변환을 한 12장의 채널 이미지들을 이용하여 계산한다. 12장의 채널 이미지는 [1]의 1~12번 이미지와 같다. 이 이미지들을 각각 그리드 분할을 하고 분할된 블록들에 대해서 평균과 분산을 구하여 학습시킨다. 서브블록은 전체이미지를 하나의 블록으로 보고 평균과 분산을 구하는 것과, 1x1에서부터 (1/5)x(1/5) 까지 서브블록으

로 나누고 각각의 블록에 대해서 평균과 분산을 구하여 이 각각을 특징점으로 저장한다. 이렇게 구해진 특징점은 총 12채널 x 55블록 x 2변수=1,320개가 된다.

#### 3.1.2. Machine Learning

이미 Flickr.com을 비롯한 많은 웹2.0 사이트에는 일반 대중의 지성을 통하여 각 이미지에 이미 수많은 태그가 첨부되어 있다. 본 논문에서는 대중의 주관이 반영되어 태깅 되어 있는 수많은 사진들을 OpenAPI 기술을 이용하여 학습 이미지로 사용하였다. 이렇게 얻어진 학습 이미지들을 앞의 특징점 추출 과정을 거쳐 특징점을 추출한 다음 이 특징점과 질의된 카테고리 정보를 이용하여 강한 분류자를 만들게 된다.

학습 알고리즘으로는 많은 방법론들이 존재하지만, 오프라인으로 수행하는 환경이므로, 속도보다는 학습의 정확도를 최대화할 수 있는 SVM 등의 방법론을 사용하는 것이 좋지만, 간단한 구현을 위하여 K-Nearest Neighbor를 이용하였다.

이렇게 시스템 학습과정을 거치게 되면 앞의 특징점 추출 과정을 통하여 얻게 된 특징들에 대한 태그를 구분할 수 있는 강력한 구분자를 얻게 된다. 이 정보를 '사진 특징-태그 DB'에 저장한다.

#### 3.1.3. Music Tagging

음악 태깅은 기존에 개발된 다양한 알고리즘들이 존재하지만, 본 논문에서는 영상의 구분에 초점을 맞추므로 음악의 경우 수동으로 구현하였다. 하지만, 추후 구현에서는 자동 음악 분류 방법론도 구현이 되어야 할 것이다.

### 3.2. Recommendation Stage

본 단계에서는, 사용자가 직접 슬라이드 쇼를 구성할 때에 '자동 음악 선곡' 메뉴를 선택함으로써 시작된다. 사용자가 슬라이드 쇼에 보여질 사진들을 선택한 후 이 메뉴를 실행하면, 시스템은 선택된 사진들을 대상으로 특징점들을 추출한다. 이렇게 추출된 특징점들을 앞에서 구축된 '사진 특징-태그 DB'에서 가장 근접한 태그를 추출한다. 각 사진들에서 많이 추출된 순서로 태그를 정렬하여 Tag Matcher에서 연관된 음악 태그들을 찾는다. Tag Matcher는 이 태그들을 이용하여 '음악-태그 DB'에서 연관된 음악들을 찾아 사용자에게 제시한다. 이 때 음악은 가장 많이 선택된 태그 순서대로 사용자에게 추천되게 된다.

#### 3.3. User Feedback Stage

사용자는 앞의 추천 단계에서 추천된 음악 리스트들에

표 2. 각 슬라이드 쇼의 선호도

	무 BGM (a)	부적절한 BGM (b)	적절한 BGM (c)
가장 좋음 (5)	2 (7%)	1 (4%)	8 (29%)
좋음 (4)	2 (7%)	6 (21%)	13 (46%)
보통 (3)	8 (29%)	11 (39%)	4 (14%)
나쁨 (2)	10 (36%)	5 (18%)	3 (11%)
가장 나쁨 (1)	6 (21%)	5 (18%)	0 (0%)
평균	2.43	2.75	3.93

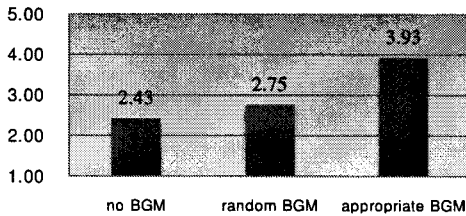


그림 2. 각 슬라이드 쇼에 대한 평균치(보통 3)

서 가장 위의 음악(즉, 가장 많은 태그가 연관된 음악)이 마음에 들지 않을 경우 아래에 있는 음악을 선택할 수 있다. 이는 시스템의 추천이 잘 못 되었음을 의미하며 이렇게 사용자에게 의해 변경된 정보는 시스템에 다시 반영되어 'Tag Matcher'에 있는 가중치를 변경하게 된다. ListenPhoto 시스템은 이러한 과정을 통하여 사용자에게 최적화된 결과를 제공할 수 있도록 한다.

#### 4. 사용자 평가

본 논문에서는 제안하는 ListenPhoto 시스템의 타당성을 증명하기 위하여 두 가지의 실험을 하였다.

첫 번째 실험에서는 ListenPhoto의 필요성을 증명하였다. 실험에는 19세부터 32세까지의 IT종사자/비종사자 총 28명을 대상으로 진행되었다. 이 실험에서는 결혼식 사진 앨범을 이용하여 배경음악이 전혀 없는 슬라이드 쇼(a)와 적절하지 않은 배경음악이 삽입된 슬라이드 쇼(b), 적절한 배경음악을 삽입한 슬라이드 쇼(c)를 각 피실험자에게 보여주고 다음의 질문을 하였다.

- 얼마나 각 슬라이드 쇼가 생생한지 (선호도)
- 각 슬라이드 쇼와 전 슬라이드 쇼를 비교하여 얼마나 느낌이 다른지 (향상도)

그 결과 표 2와 3은 각 슬라이드 쇼에 대한 사용자들의 호감도와 (a)와 (b), (b)와 (c)를 비교했을 때의 향상도를 보여주며 그림 2와 3은 이에 대한 그래프 결과를 보여준다. 이는 다음의 두 가지 사실을 보여준다.

- 사진만 단독으로 볼 때보다 더 풍부한 자극(특히 음악적 자극)이 가미될 때 감상에서 감동을 더해줄 수 있다.
- 적절한 배경음악의 제공이 더 좋은 느낌을 준다.

표 3. 두 슬라이드 쇼간 향상도

	무 BGM (a) 보다 부적절한 BGM (b) 이		부적절한 BGM (b)보다 적절한 BGM (c) 이	
훨씬 좋아짐	2	7%	10	36%
좋아짐	11	39%	16	57%
그대로	7	25%	2	7%
나빠짐	5	18%	0	0%
훨씬 나빠짐	3	11%	0	0%

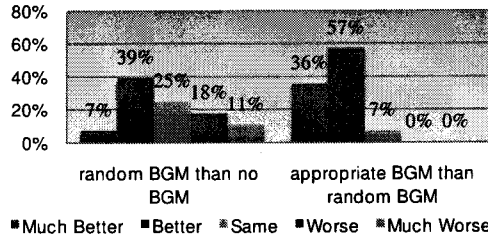


그림 3. 두 슬라이드 쇼간 비교

특히, 슬라이드 쇼(b)에서는 (a)에 비하여 비슷하거나 더 나빠졌다는 의견도 54%를 차지했다. 이는 기존 시스템과 같은 임의의 음악의 선택이 오히려 보는 이에게는 감상을 저해할 수 있음을 보여준다.

다음 실험은 ListenPhoto의 성능을 측정하는 실험을 진행하였다. 이 프로젝트에 참여하지 않는 4명의 동료를 대상으로 ListenPhoto 시스템과 구글의 Picasa를 직접 사용해보고 이에 대한 만족도를 평가하였다. 이 실험은 제안하는 컨셉의 실현 가능성을 검증하기 위함으로 간단하게 시스템을 구성하였다. - 3개 카테고리(결혼, 여름 휴가, 졸업)에 대해 각각 20장씩의 트레이닝 이미지를 이용해 학습하였으며, 이미 태깅된 100개의 음악을 이용하였다. 감상에는 각 카테고리당 10장씩 2세트의 앨범을 감상하도록 하였다. (총 6 앨범)

전체적으로 사용자들의 반응은 긍정적이었다. 사용자들은 Picasa에 비하여 제안하는 시스템이 더 낫다는 답변을 하였다. Picasa의 경우, 슬라이드 쇼가 바뀌어도 단일한 배경음악만 나온 반면, 제안하는 시스템은 감상하는 슬라이드 쇼에 맞춰 적절하게 배경음악을 추천함으로써 사용자들의 만족도를 높였다. 하지만, 아직 완벽하게 분석하지 못하여 간혹 잘못된 음악을 추천하는 경우도 있었다. 사용자들은 이에 대하여 좀 더 신뢰성 있는 시스템의 구축이 필요할 것이라는 답변을 보여주었다.

#### 5. 결론

본 논문에서는 영상 분석을 이용해 사진 앨범을 감상하는 사용자에게 적절한 배경음악을 제공함으로써 풍부한

자극을 제공하는 ListenPhoto 시스템을 제안하였다. 이 시스템은 태터툴즈 등의 블로그나 싸이월드나 Facebook 등의 SNS(Social Networking Service) 시장이나 Flickr, Fotki 등 디지털 사진 관리 시장 또는 물리적인 전자 앨범, DTV 등에서 크게 사용될 가능성이 있다. 하지만, 앞으로 좀 더 정확한 특징 추출 알고리즘 구현, 실시간 응답이 가능한 적합한 특징점-태그 DB 구성, RM-RLS, OS-ELM 등 재귀적인 구분자 구현, 자동 음악 구분 알고리즘 구현 등에 대한 보충이 필요할 것이다.

## 6. 감사의 글

본 연구는 교육인적자원부의 두뇌한국 21(BK21), 한국과학기술재단의 특정기초연구(R01-2005-000-10898-0), LG 전자 DTV 연구소와의 BK21 사업체 대응 프로젝트(2007-8-0488)의 지원으로 이루어진 연구임을 밝힙니다.

## 7. 참고문헌

- [1] Baluja, S. Rowley, H.A. Large scale performance measurement of content-based automated image-orientation detection. In *Proceedings of the Image Processing, 2005. ICIP 2005*. IEEE, Vol 2. pp.514-517.
- [2] Chen, J., Chu, W., Kuo, J., Weng, C., and Wu, J. 2006. Audiovisual slideshow: present your journey by photos. In *Proceedings of the 14th Annual ACM international Conference on Multimedia. MULTIMEDIA '06*. ACM Press, New York, NY, 955-956.
- [3] Chen, J., Chu, W., Kuo, J., Weng, C., and Wu, J. 2006. Tiling slideshow. In *Proceedings of the 14th Annual ACM international Conference on Multimedia. MULTIMEDIA '06*. ACM Press, New York, NY, 25-34.
- [4] Hua, X.-S. Lu, L. Zhang, H.-J. Content based Photo-graph Slide Show with Incidental Music. In *Circuits and Systems, 2003. ISCAS '03. Proceedings of the 2003 International Symposium on*, ACM, 2003, pp. II-648 - II-651 vol.2.
- [5] Shamma, D. A., Pardo, B., and Hammond, K. J. 2005. MusicStory: a personalized music video creator. In *Proceedings of the 13th Annual ACM international Conference on Multimedia. MULTIMEDIA '05*. ACM Press, New York, NY, 563-566.