

특집논문 (Special Paper)

방송공학회논문지 제21권 제4호, 2016년 7월 (JBE Vol. 21, No. 4, July 2016)

<http://dx.doi.org/10.5909/JBE.2016.21.4.525>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

스토리보드에 따라 장면요소를 자동 조합하는 주제모델링 기반 온라인 비디오 매쉬업 시스템 개발

박종빈^{a)†}, 김경원^{a)}, 정종진^{a)}, 임태범^{a)}

Development of Online Video Mash-up System based on Automatic Scene Elements Composition using Storyboard

Jongbin Park^{a)†}, Kyung-Won Kim^{a)}, Jong-Jin Jung^{a)}, and Tae-Beom Lim^{a)}

요 약

본 논문은 제한된 개수의 장면요소들을 주어진 스토리보드에 따라 자동으로 선택 및 배치하여 다양한 목적과 의도를 반영하는 동영상 생성하는 방법을 제안한다. 동영상 제작하려는 사용자가 사진이나 동영상을 온라인 시스템 상에 업로드 하면 미리 확보된 메타 정보들을 활용해서 장면요소에 해당하는 장면템플릿들을 선택 및 배치하고, 내부 리소스를 수정하며, 메타정보에 부합하는 필터효과나 배경음악을 포함하도록 한다. 이는 입력 사진들을 슬라이드 재생방식으로 구성하는 비디오 콜라주 기법이나 템플릿 기반 리소스 치환 방법들이 지닌 단점들을 보완하면서 장점을 더욱 강화할 수 있는 방법이라 할 수 있다. 제안 방법은 마케팅 및 홍보전략 의도 등을 이용하여 홍보 동영상을 쉽게 제작하는 온라인 서비스 시스템에 적용하여 실 서비스 활용 가능성을 검증하였다.

Abstract

In this paper, we develop an online video mash-up system which use automatic scene elements composition scheme using a storyboard. There are two conventional online video production schemes. Video collage method is simple and easy, but it was difficult to reflect narrative or story. Another way is a template based method which usually select a template and it replaces resources such as photos or videos in the template. However, if the related templates do not exist, there are limitations that cannot create the desired output. In addition, the quality and atmosphere of the output is too dependent on the template. To solve these problems, we propose a video mash-up scheme using storyboard and we also implement a classification and recommendation scheme based on topic modeling.

Keyword : Storyboard, Video Mash-up, Scene composition, Topic Modeling, Template

a) 전자부품연구원(Korea Electronics Technology Institute)

† Corresponding Author : 박종빈(Jongbin Park)

E-mail: jpark@keti.re.kr

Tel: +82-2-6388-6699

ORCID: <http://orcid.org/0000-0002-6123-5300>

※ 본 연구는 미래창조과학부 및 정보통신기술진흥센터(IITP)의 SW컴퓨팅산업원천기술개발사업의 일환으로 수행하였음. [과제번호(B0101-16-0559), 디지털 소상공인 지원을 위한 지역 비즈니스 전략 분석 및 맞춤형 영상홍보 창작 SW 플랫폼 개발]

※ 이 논문의 연구결과 중 일부는 “2016년 IWAIT 2016 국제학술대회”에서 발표한 바 있음.

· Manuscript received March 23, 2016; Revised June 10, 2016, Accepted June 10, 2016.

I. 서론

동영상은 방송미디어와 인터넷 기반 스트리밍 환경에서 광고, 교육, 전시, 기록, 의사소통 등과 연계한 다양한 응용서비스들을 제공하기 위한 핵심요소이다^[1-3]. 동영상 콘텐츠 생성관점에서는 최근의 기술·환경적 진화와 더불어 전문가가 아닌 일반인들도 직접 촬영한 사진이나 동영상으로 개성 있는 동영상들을 만들고 공유하고자 하는 니즈가 꾸준히 증가하고 있다. 따라서 동영상을 쉽고 빠르게 제작할 수 있도록 하는 편집 도구나 서비스들 역시 다양하게 제시되고 있다. 대표적으로 비디오 콜라주 방식이 있으며, 이는 사진이나 영상을 한꺼번에 선택하고 원하는 분위기를 선택하면 입력된 콘텐츠들을 하나로 엮어서 동영상을 생성 하는 방식이다^[4]. 그러나 비디오 콜라주는 입력 콘텐츠에 필터효과(Filter effect)나 전환효과(Transition effect)를 부가하여 결과물을 합성 하므로 기존의 이미지 뷰어들이 제공하던 이미지 슬라이드 재생 형태에서 크게 벗어나지 못하는 한계도 존재했다. 게다가 이 방식은 리소스를 시간 순서대로 나열하고 국부적으로 영상 효과를 추가하는 원리이므로 동영상 제작자의 의도나 이야기구조를 반영할 수 있는 방식은 아니다^[4].

쉬운 동영상 제작을 가능케 하는 또 다른 방법으로써 템플릿 기반 방법이 대표적이다^[5-8]. 이 방법은 사전에 디자인된 템플릿을 선택하고 원하는 미디어를 대체·삽입하는 방식으로 영상 제작이 진행된다. 완성도 높은 템플릿을 일단 만들어 놓기만 하면 사진이나 동영상만 대체하여 여러차례 재활용이 가능하므로 돌잔치 영상이나 성장 동영상을 제작해주는 다수의 업체들이 이 방식을 주로 사용한다^[5]. 하지만 템플릿 기반 방법은 결과물의 분위기나 특징이 템플릿에 크게 의존하여 서로 다른 사용자가 같은 템플릿을 사용하면 결과물 간의 차이점이 부각되기 어려워지는 문제점이 있었다. 개성 있고 차별화된 동영상 제작을 원하는 사용자들에게는 단점으로 지적될 수 있는 부분이다. 이를 해결하기 위해서는 가능한 많은 수의 템플릿들을 제작해야 하는데 이는 서비스 준비 비용의 증가로 이어질 수 있다.

따라서 본 논문에서는 사용의 편의성과 결과물의 다양성을 동시에 제공하는 방법으로써 주어진 스토리보드에 따라 장면템플릿이나 리소스와 같은 다양한 장면요소들을 서로

조합하여 영상을 합성하는 방식을 제안한다. 이때 분류와 추천을 위해 장면요소들에는 메타정보가 포함됨을 가정하며 메타정보를 LSI(Latent Semantic Indexing) 방식의 주제 모델링 방법을 사용하여 분류하고 추천한다. 최종 합성될 영상은 전략메타정보로 불리는 일종의 사용자 컨텍스트(User context) 정보를 입력받아, 주제모델링 방식으로 구조화된 템플릿들을 조합하여 전체 영상을 합성할 수 있는 프로젝트 파일을 자동으로 생성한다. 영상 합성에 사용될 사진이나 동영상은 가이드 UI를 통해 사용자에게 직접 요청하거나 이 과정이 생략되면 메타정보를 사용해서 자동으로 채워 넣는 방식을 사용한다. 결과적으로 본 논문의 제안방법을 통해 제한된 개수의 템플릿들을 가지고서 스토리보드에 따른 조합 및 변형을 통해 다양한 목적과 의도에 부합하는 동영상 콘텐츠를 만들 수 있게 된다. 이는 장면템플릿과 같은 전문적인 디자인 작업 결과를 재활용 할 수 있으면서도 상위 수준의 이야기 구조를 적용할 수 있게 되어 개성 있고 차별화된 동영상 생성을 가능하도록 한다.

본 논문의 각 장별 구성 내용은 II장에서는 제안방법의 기반이 되는 기존의 연구내용들을 소개하고, III장에서는 논문에서 제안하는 비디오 매쉬업 방법을 설명한다. IV장에서는 구현내용을 보이며, V장에서 결론 및 향후 개발 내용을 정리한다.

II. 기존 연구 내용

1. 기존의 온라인 비디오 제작 서비스 시스템

그림 1은 [8]에서 제시한 기존의 온라인 비디오 제작 서비스 시스템에 대한 개념도를 나타낸다. 동영상을 제작하기 위해서 각종 메타정보, 사진, 동영상과 같은 리소스를 입력받고 이를 전략메타정보에 따라 선별하여 최종적으로 비디오를 출력하는 구성이다. 기존의 웹 기반 동영상 제작 서비스들과의 차이점은 메타정보를 이용하여 제한된 템플릿들을 분류하고 추천하여 의도에 맞는 콘텐츠를 제작할 수 있도록 하는 것이다. 여기서 메타정보는 정형화된 형태이거나 정해진 형태가 없는 일상적인 텍스트 문자열 포맷을 가정한다^[8].

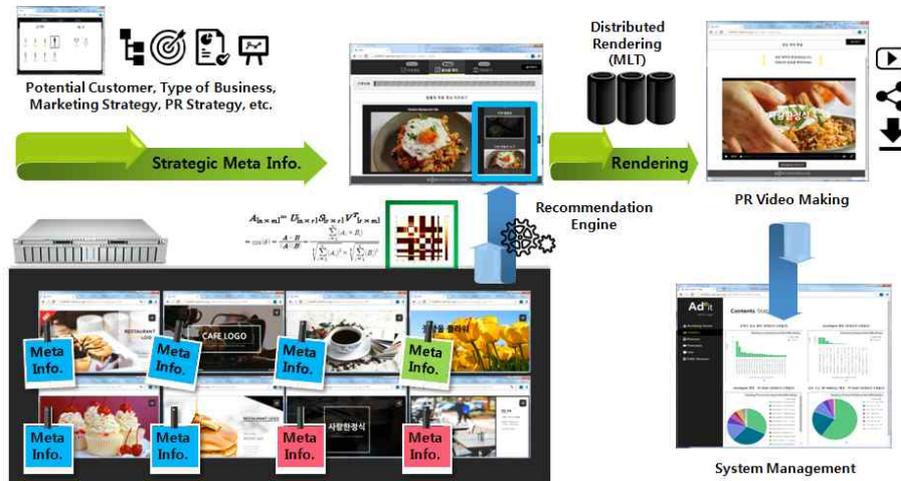


그림 1. 기존의 온라인 비디오 제작 프레임워크 개념도^[8]
 Fig. 1. Concept of previous online video making framework^[8]

본 논문에서는 [8]과는 달리 각종 장면요소들을 단순 선택하는데 그치지 않고, 상호 조합하거나 필터 등을 추가하여 제한된 개수의 템플릿들을 사용하더라도 다양한 스토리 구성과 분위기를 연출할 수 있도록 하는 방법을 제안한다.

2. 동영상 제작을 위한 플롯과 스토리보드의 정의

표준국어대사전에서는 용어 “플롯”을 “몇 가지 부분이나 요소들을 모아서 일정한 전체를 짜 이룸. 또는 그 이룬 결과”, 문학적 분야로 한정하여 “문학 작품에서 형상화를 위한 여러 요소들을 유기적으로 배열하거나 서술하는 일”, 미술 분야에 한정하여 “색채와 형태 따위의 요소를 조화롭게 조합하는 일”로 정의하고 있다^[9]. 위키백과 사전에서는 플롯을 네러티브라고도 부르며 상위 수준의 이야기 구조라고 정의하고 있다^[10]. 이러한 플롯을 시간순서에 따라 더욱 구체적으로 표현한 것이 스토리이다^[11,12]. 스토리를 구체적으로 표현하는 방법으로써 스토리보드 있다. 스토리보드는 동영상 제작 시 전체적인 촬영 및 편집 계획과 연관되며, 통상적으로 최종 영상물을 만들기 전에 작성한다^[11,12]. 이러한 스토리보드는 영상의 흐름이나 구성요소, 촬영정보 등을 미리 약속된 기호나 그림으로 스케치하여 작성하는 것이 일반적이다^[11,12].

본 논문에서는 플롯과 스토리보드 개념을 동영상 생성

시 활용하기 위하여 장면별로 정형메타정보와 비정형메타 정보를 사용하는 방법을 제안한다. 이를 통해 사진이나 리소스들을 단순 나열하는 방식에서 더욱 진화하여 작가나 디자이너들이 구상한 추상적인 이야기 구조를 동영상에 반영할 수 있도록 지원한다.

3. MLT 멀티미디어 프레임워크

MLT(Media Lovin' Toolkit) 멀티미디어 프레임워크는 다양한 미디어 소스들을 시간에 따라 배치하고 전환 및 필터효과, 2차원 및 3차원 그래픽 효과를 부가하여 실제 영상을 만들 수 있도록 지원한다^[13]. MLT는 공개소스 멀티미디어 프로젝트로써 방송용 콘텐츠를 비롯한 다양한 비디오 콘텐츠 창작을 위해 개발되고 있다^[13]. 외부 프로그램들과의 연동을 위한 API가 제공되어 다양한 프로그래밍 언어로 접근이 가능하며 방송 콘텐츠, 비디오 편집기, 미디어 재생기, 변환부호화기, 비디오 스트리밍과 같은 여러 응용에 활용되고 있다^[13]. 게다가 MLT는 LGPL 라이선스를 따르므로 상용솔루션 개발에도 활용 가능하다. 기능 측면에서는 멀티미디어 콘텐츠들을 쉽게 조합하고 표시하는 필수 기능들을 제공하고, 플러그인 구조를 도입하여 지속적인 기능 확장을 가능하게 한 것도 주요 장점이다^[13]. 단, 시스템 관점에서 MLT는 영상생성에 필요한 저수준 기능들에 특화

되어 있으므로 주어진 스토리보드에 따른 영상 합성을 위해서는 부가적인 정보처리 기술이 요구된다.

4. 주제모델링 기반 장면요소 분류 및 추천

기존 연구내용으로써 [8]에서는 동영상 제작 시 전략메타정보로 정의되는 각종 부가정보를 기반으로 영상 합성에 필요한 장면템플릿들을 분류하고 추천하는 방법을 제안했다. 이때 N 개의 장면템플릿(Scene template)이 식 (1)과 같이 마련되어 있다고 가정한다. 여기서 N 은 유한한 정수 값을 의미하며, 장면템플릿 각각은 식 (2)와 같은 메타정보들을 가지고 있음을 가정한다. 그리고 메타정보의 형태는 텍스트 문자열(Text String)의 집합으로 이미 주어졌음을 가정한다.

$$S = \{S_1, S_2, \dots, S_N\} \quad (1)$$

$$D = \{D_1, D_2, \dots, D_N\} \quad (2)$$

각각의 장면템플릿들은 식 (3)과 같은 사진, 동영상, 음악, 텍스트정보 등의 리소스들로 구성된다.

$$R = \{R_1, R_2, \dots, R_N\} \quad (3)$$

장면템플릿들은 식 (4)와 같이 사용자들이 제공한 사진이나 동영상 혹은 음악 파일들을 이용해서 식 (5)와 같은 프로젝트 파일을 구성한다. 임의의 프로젝트 파일은 MLT 프레임워크에서 해석 가능한 XML 파일 정보와 영상을 만들 때 필요한 사진이나 동영상 등을 모두 포함하게 된다. 분위기나 느낌을 결정하는데 장면템플릿 자체의 영향이 매우 크기 때문에 적절한 장면템플릿, 내부 리소스나 필터효과 선택과 조합이 중요하다. 이런 특징으로 인해 본 논문에서는 [8]에서 주제모델링 방식으로 장면템플릿을 선택하던 방식을 확장하여 식 (1), 식 (3), 식 (4)와 같은 장면요소들의 상호 조합에 의해 통합된 이야기를 구성 하도록 한다.

$$R' = \{R'_1, R'_2, \dots, R'_N\} \quad (4)$$

$$P = \{P_1, P_2, \dots, P_N\} \quad (5)$$

장면템플릿을 추천하기 위해 [8]에서는 LSA(Latent Semantic Analysis/Indexing)^[14] 방법을 사용했다. LSA 방법은 입력된 문서들을 벡터공간(Vector Space)에 표현하고 SVD(Singular Value Decomposition)를 수행하여 주제(Topic)들을 찾고, 이를 통해 벡터공간에서 문서들을 표현하는 차원을 줄이거나, 유사도를 분석하여 주제별로 분류 및 추천하는 기능을 제공한다^[14,15]. 주제모델링이 완료된 상태에서 각종 전략메타정보에 따라 장면템플릿을 추천하기 위해서는 유사도 분석이 필요하다. 두 벡터 A , B 사이의 유사도(similarity)를 계산하기 위해서는 식 (6)과 같은 코사인 유사도(Cosine similarity) 공식을 사용한다. 즉, 유사도가 -1인 경우 완전히 다름, 0인 경우 상관성 없음, 1인 경우 완전히 일치함을 의미한다.

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n (A_i \times B_i)}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (6)$$

III. 제안방법

1. 제안하는 비디오 매쉬업 방법의 개요

그림 2는 제안하는 비디오 매쉬업 방법의 순서도이다. 비디오 매쉬업을 위해서는 사용자 정보, 사용자 의도 정보, 텍스트 형태의 메타정보로 기술되는 전략메타정보들이 시스템에 사전에 확보되어 있음을 가정한다^[8]. 전략메타정보는 다양한 경로 및 방법으로 수집될 수 있지만^[8] 본 논문에서는 홍보용 비디오 매쉬업이라는 실시 예를 가정하여 업종 정보, 마케팅전략, 위치정보, 분위기정보, 사용자 의도정보, 홍보 시간대 정보로 한정한다. 예를 들어 업종은 “커피숍”, 마케팅 전략은 “20대 여성들이 선호하는 홍보물을 정오에 배포할 것”, 사용자 의도는 “아메리카노를 중점적으로 광고 원함”과 같은 전략메타정보가 제시될 수 있다. 이후 메타정보를 나타내는 문자열들에서 사전(dictionary)에 정의된 명사(none) 위주로 키워드를 추출하고 [8]의 주제모델링 방법을 써서 장면 조합을 위한 유사도 계산에 사용한다. 단, 형태소 분석 및 처리에 따라 알고리즘의 변형이나 추가 성능 향

상이 가능하겠지만 이는 본 논문의 범위를 벗어난다.

그림 2의 스토리보드(Storyboard)는 작가나 디자이너 등에 의해 만들어진 일종의 이야기 구조로써 사전에 확보되어 있고 사용자는 UI를 통해 원하는 분위기나 테마를 갖는 스토리보드를 미리 선택했다고 가정한다. 이러한 스토리보드의 표현을 위해 본 논문에서는 여러 개의 장면이 스토리보드를 이룬다고 가정하고 장면에 따라 장면번호, 장면스

```

{
  number_of_scenes: "3",
  - scenes: [
    - {
      scene_id: "1",
      sketch_path: "/img/1.png",
      RegularTags: "Opening",
      IrregularTags: "Cafe,Coffee,Americano"
    },
    - {
      scene_id: "2",
      sketch_path: "/img/2.png",
      RegularTags: "MainBody",
      IrregularTags: "Menu,LowPrice"
    },
    - {
      scene_id: "3",
      sketch_path: "/img/3.png",
      RegularTags: "Closing",
      IrregularTags: "Title,Address"
    }
  ]
}
    
```

그림 2. 스토리보드를 표현하는 JSON 형식의 데이터 예시
 Fig. 2. Example of a JSON format data for storyboard representation

캐치, 정형태그, 비정형태그를 JSON(JavaScript Object Notation) 포맷으로 기술하여 상위수준의 이야기 구조를 만들 수 있도록 했다. 예를 들어 그림 2는 스토리보드를 표현하는 JSON 형식의 데이터 예시이다.

본 논문에서 사용하는 추천방법을 변형하면 원하는 스토리보드를 자동으로 추천하는 것도 가능하지만 이는 논문의 범위를 벗어난다. 이러한 가정 하에 그림 3에 나타낸 절차에 따라 비디오 합성이 진행된다. 그림 3은 스토리보드에 따라 장면템플릿(Scene Template), 리소스(Resource), 배경음악(Background Music), 필터효과(Filter Effect), 마스크효과(Mask Effect)를 선택하는 과정이 진행되어 최종적으로 비디오를 출력하는 과정을 보여준다. 본 논문에서는 이러한 5가지 구성 요소를 장면요소로 칭하며 이들의 선택과 조합을 통해 동영상상을 생성한다. 이는 스토리보드에 따른 이야기 구조를 반영하되 - (1) 전문 디자이너에 의해 개발된 템플릿을 사용할 수 있도록 할 것, (2) 사용자 별로 입력한 리소스들을 반영할 것, (3) 전략메타정보에 의한 차별화된 분위기나 효과를 지원할 것 - 이라는 자체 요구사항을 달성하기 위함이다.

2. 스토리보드에 따른 자동 장면 구성방법

본 절에서는 입력된 스토리보드와 전략메타정보에 맞춰 장면템플릿, 리소스, 배경음악, 필터효과, 마스크효과와 같

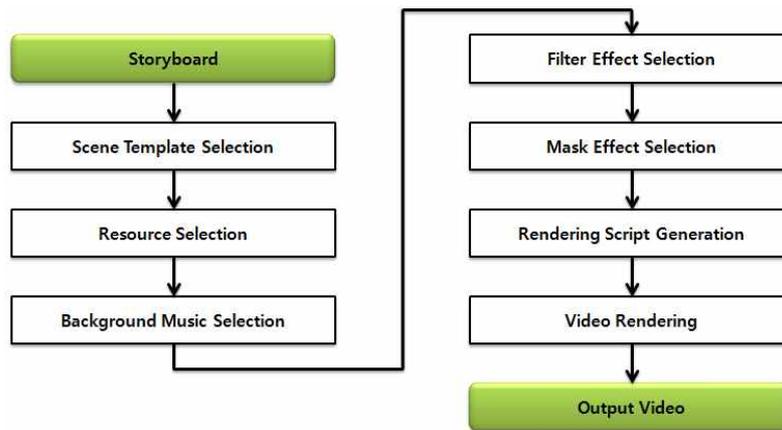


그림 3. 제안하는 비디오 매쉬업 방법의 순서도
 Fig. 3. Flowchart of proposed video mash-up scheme

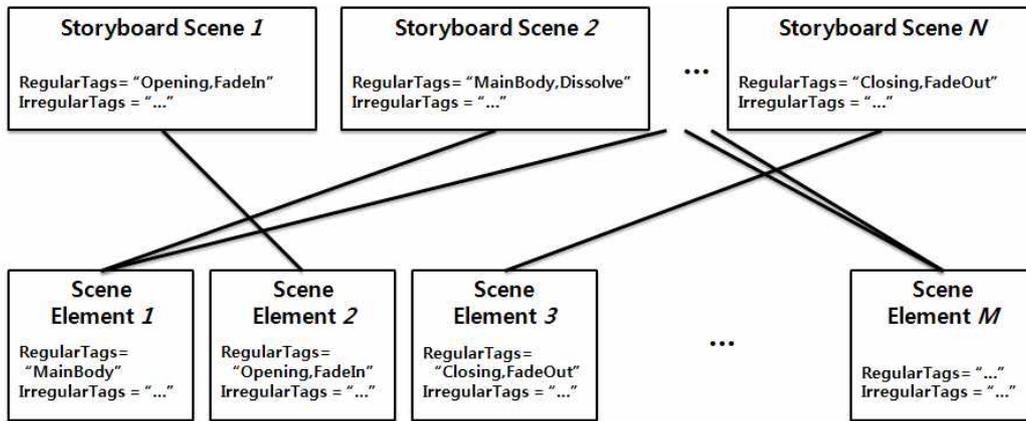


그림 4. 스토리보드와 장면요소들 사이의 연결 알고리즘의 예
 Fig. 4. Example of Interface between storyboard and scene elements

은 장면구성요소들을 자동으로 선택하고 조합하는 상세한 방법을 기술한다. 이를 위해 스토리보드와 장면구성요소들 사이에 메타정보 기반의 연결 인터페이스를 정의하고, 상호 연결시킬 수 있는 알고리즘을 제안한다. 그림 4는 이에 대한 상세 설명을 위한 예시이다. 그림 4에서 하나의 스토리보드는 N 개의 장면(Scene)으로 구성되고, 장면템플릿, 리소스, 각종효과 등의 장면요소들이 각각 M 개씩 준비되어 있다고 가정한다. 스토리보드를 구성하는 각각의 장면(Storyboard Scene)들과 장면요소(Scene Element)들은 각각 정형메타태그들(RegularTags)과 비정형메타태그들(IrregularTags)을 포함한다. 여기서 정형메타태그들은 사전에 스토리보드와 장면요소를 연계하기 위해 정의한 전체 메타태그집합의 부분집합으로써 플롯 혹은 스토리구조에 맞게 장면요소들을 추출하기 위해 사용한다. 비정형메타태그들은 사전에 정의되지 않고 동적으로 갱신되는 사전(Dictionary)에 포함된 단어들의 집합으로써 [8]의 방법으로 처리된다. 본 논문에서는 비정형메타태그들을 기반으로 주제모델링 방법을 사용하여 스토리보드의 장면과 장면요소들 사이의 유사성을 계산한다.

그림 4의 연관 방식에 따라 스토리보드가 주어졌을 때 스토리보드 각각의 장면에 맞는 장면요소들을 선택함으로써 추상적인 스토리보드가 동영상 합성을 위한 구체적인 프로젝트 파일로 변환된다. 예를 들어 그림 4에서 첫 번째 스토리보드의 정형메타태그들은 “Opening, FadeIn”으로 표

시되어 있다. 따라서 M 개의 장면요소들의 정형메타태그들을 검사하여 “Opening” 혹은 “FadeIn”이 전체 혹은 부분적으로 검색되는 경우 후보 장면요소로 추출될 수 있다. 이렇게 추출된 후보 장면요소들에 대해서 [8]의 주제모델링 방식에 따른 유사도 계산을 수행하여 가장 근사한 장면요소를 선정한다. 그림 5는 이에 대한 개념도를 나타낸 것으로써 사전에 마련된 특정 장면요소의 전체 집합(Entire Set of Specific Scene Elements)에서 정형메타태그들의 검색을 통해 부분집합이 선택되고, 마지막으로 비정형메타태그들을 통해 유사도 계산이 완료된 최종 장면요소의 부분집합을 추출한다. 정형메타태그들의 경우는 단순 단어 검색을 통

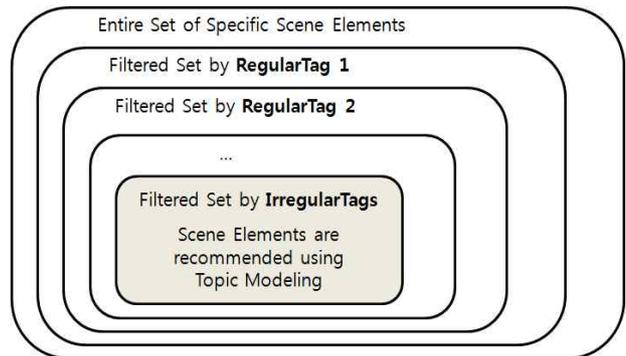


그림 5. 특정 장면요소의 전체 집합에서 정형메타태그와 비정형태그를 사용하여 장면요소의 부분집합을 선택하는 예시
 Fig. 5. Example of scene elements extraction using RegularTags and IrregularTags

해 간단하게 추출이 가능하지만 비정형메타태그들은 여러 단어들의 가중치를 종합적으로 반영할 수 있도록 하기 위해 본 논문에서는 LSA방법을 사용한다. 하지만 응용에 따라 TF-IDF방법이나 LDA와 같은 주제모델링 방법을 사용할 수 있다^[4-16].

이 과정을 본 논문의 5가지 장면요소에 대해 적용해 보면 스토리보드에 포함된 특정 장면을 결정하기 위해 장면요소 중에서 장면템플릿을 정형메타태그와 비정형메타태그를 통해 선정한다. 이후 해당 장면에 들어갈 리소스, 배경음악, 필터효과, 마스크효과를 동일 방법으로 선정한다. 이 과정에 대한 구체적인 예시를 그림 6에 나타내었다.

그림 6은 그림 2와 같은 스토리보드에 따라 장면요소들 중에서 장면템플릿을 선택하고, 동영상 합성을 위해 비디오 트랙을 구성하는 방법의 예시로서, rTag는 정형메타태그, iTag는 비정형메타태그를 의미한다. 그림 6의 예시에서 주어진 스토리보드의 장면별로 정형메타태그만을 추출한 목록은 {rTag:“Opening”, rTag:“MainBody”, ..., rTag:“MainBody”, rTag:“Ending”, rTag:“BGM”, rTag:“Filter”, rTag:“Mask”} 임을 가정하고 있다. 첫 번째 스토리보드 장면의 rTag가 “Opening”이므로 장면템플릿들의

rTag가 “Opening”인 것들만 먼저 추출된다. 그리고 스토리보드의 iTag와 장면템플릿의 iTag 사이에 주제모델링 기반 유사도 값을 식(6)을 사용하여 계산하고 유사도 값이 가장 높은 장면템플릿을 “Opening” 장면에 사용할 장면템플릿으로 선정한다. “MainBody”, “Ending”에 대해서도 마찬가지로 절차를 따른다. 도입과 결말템플릿들의 내용은 영상을 시작하거나 마무리 짓기 위한 효과나 글을 삽입하는 것이 일반적이다. 메인바디를 위한 정형태그들로는 액션, 잔잔함, 풍경, 클로즈업, 패닝, 줌아웃, 제품소개, 가격소개, 위치 정보 등의 분위기나 내용이 포함된 형태로 분할되고 구성될 수 있다.

장면요소 중에서 배경음악, 필터효과, 마스크효과는 동영상 분위기의 일관성을 위해 동영상이 재생되는 전체 시간에 대해 전역적으로 반영되도록 구성했다. 예를 들어 그림 6에서는 스토리보드 장면에 포함된 rTag가 “BGM”, “Filter”, “Mask”를 가지는 경우로써 음악이나 필터효과는 장르, 내용, 분위기에 따라 다르게 선택되어야 하는바, iTag를 사용하여 주제모델링 방식으로 음악이나 효과를 선택한다.

결과적으로 장면템플릿들은 크게 도입(Opening)템플릿, 메인바디(Main Body)템플릿, 결말(Ending)템플릿 순서로

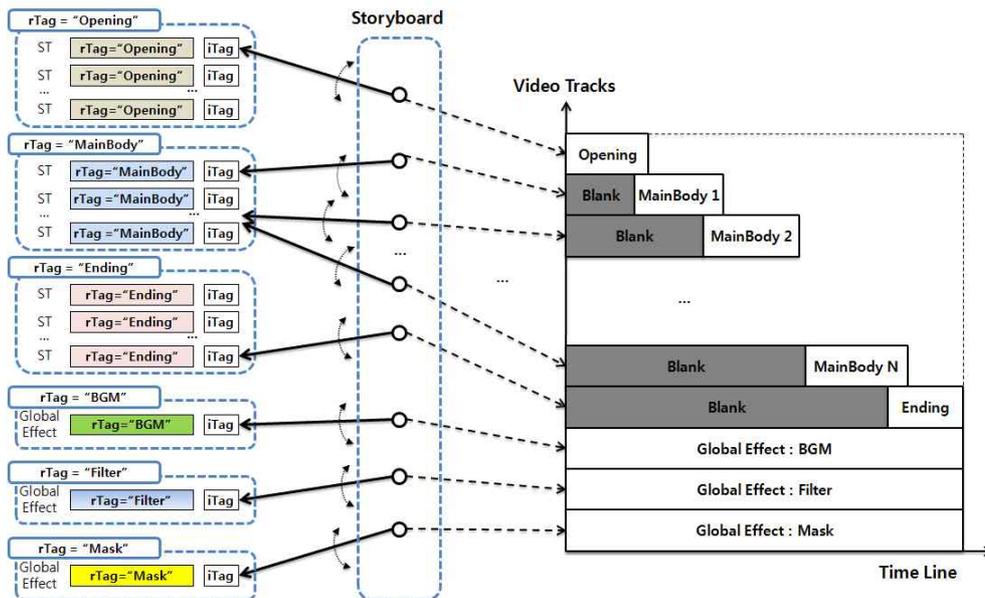


그림 6. 장면템플릿 선택과 비디오 트랙 구성 방법의 예
 Fig. 6. Example of Selection of Scene Templates and Composition of Video Tracks

배치된다. 본 논문에서는 장면템플릿은 대개 10초 이내로 구성했으며 이를 조합하여 한편의 동영상을 완성하도록 했다.

IV. 개발결과

1. 시스템 개발 내용 요약

그림 7은 제안 서비스를 위해 개발한 사용자환경, 소프트웨어 및 하드웨어 시스템구성을 나타낸다. 물리적으로는 (1) 프론트페이지와 데이터베이스 서버, (2) 분산 렌더러 서버, (3) 콘텐츠 관리 서버로 분리 구성된다.

먼저 프론트페이지(Front-page) 서버는 각종 브라우저 혹은 모바일 앱에서 접속 시 UI(User Interface)페이지 제공을 담당하고 동영상 제작과 관련된 데이터를 통합적으로 관리하는 역할을 수행한다. 서비스 제공을 위해 아파치 웹 서버가 설치되었고, PHP언어와 MariaDB가 연동되어 웹 서비스가 구현된다. 제안방법으로 구성된 영상을 실제 동영상으로 만드는 부분은 분산 렌더러 서버가 담당하여 프론트 페이지 서버와 연계하여 렌더링 작업 큐(Rendering Queue)에 렌더링 작업(Job)이 발생하면 휴면상태(Idle State)인 렌더러가 최우선으로 할당되어 작업을 수행하는 방식으로 병

렬 렌더링을 수행한다. 이는 동시 렌더링 처리 수를 높이기 위해 고안된 것으로써, 제안 시스템의 렌더러는 3GHz 클럭의 CPU, 8GByte 램을 가진 하드웨어 구성에서 약 30초 분량의 동영상을 렌더링 하기 위해 약 3분~6분 정도의 시간이 소요되는데 렌더러 서버를 늘일수록 선형적으로 렌더링 대기시간을 줄일 수 있게 된다.

그리고 CMS(Contents Management Server)는 각종 리소스 파일들과 만들어진 영상을 저장하고 스트리밍 하는 역할을 수행한다.

그림 8은 제안방법을 사용해서 장면요소들을 조합하여 결과를 JSON 형태로 표시한 예시이며 시작장면과 종료장면 사이에 5개의 메인바디 장면템플릿으로 구성 되었다. 장면템플릿은 MLT 프레임워크에서 해석될 수 있는 XML 스키마를 따르며, 장면템플릿별로 사진, 동영상, 텍스트와 같은 리소스가 포함되는데 자동으로 리소스를 선택해야 하는 경우 사용자가 업로드한 리소스 중에서 정형메타태그와 비정형메타태그 매칭을 수행하여 유사도가 가장 높은 리소스를 선택한다. 그림 8의 아래쪽에는 “bgm”, “filters”라는 항목이 존재하여 원하는 배경음악, 각종 필터나 마스크효과를 추가할 수 있도록 했다. 배경음악은 동영상 1개당 하나의 항목으로 제한했으며, 필터효과는 복수 개를 사용해도 무방하다. 필터효과의 예로는 “그레이스케일(Grayscale)”, “올드필름(Oldfilm)”, “세피아(Sepia)” 등이 있으며 MLT에

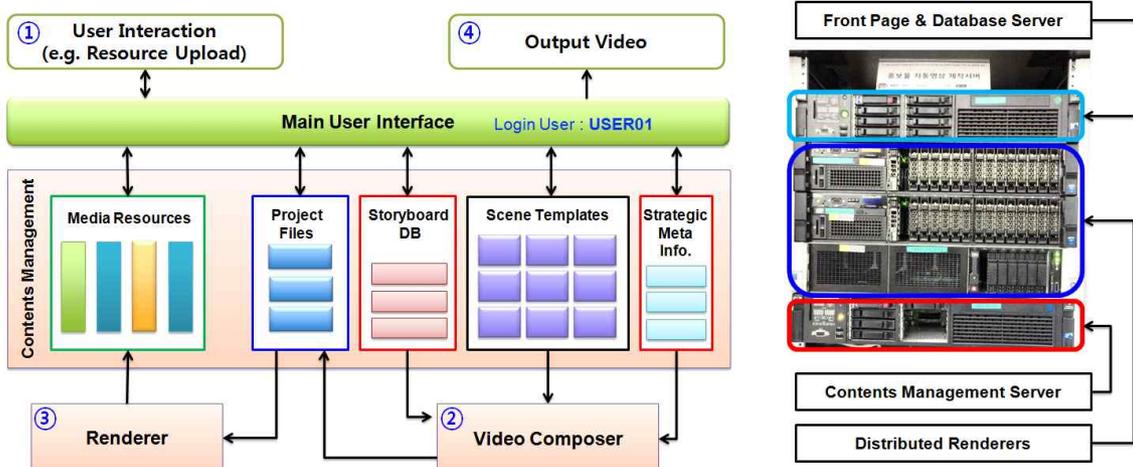


그림 7. 제안 서비스를 위한 시스템 구성
Fig. 7. System Configuration for the Proposed Service

```

{
  "mltfiles":
  [
    "xml/plot_intro_4tracks.mlt",
    "xml/plot_mainbody_menu1.mlt",
    "xml/plot_mainbody_menu2.mlt",
    "xml/plot_mainbody_interior1.mlt",
    "xml/plot_mainbody_interior2.mlt",
    "xml/plot_mainbody_location.mlt",
    "xml/plot_ending_1.mlt"
  ],
  "bgm": "effects/audio_xmas_jinglebells.xml",
  "filters":
  [
    "effects/mask_xmas.xml",
    "effects/mask_made_by_keti.xml"
  ]
}
    
```

}. Composition of scene templates

}. Background Music

}. Post Filter

}. Overlay Mask

}. Watermark

그림 8. 장면템플릿을 조합하기 위한 JSON 파일의 예
 Fig. 8. Example of a JSON file for scene templates composition

서 제공하는 약 100개 이상의 기본 필터를 서로 조합하고 매개변수를 수정하여 사용할 수 있다¹³⁾.

그림 9는 이와 같은 원리를 반영하도록 핵심 항목만으로 구성된 MLT 파일의 예이다. 사진 1장으로 구성된 장면 2개

가 시간 순으로 배열되었으며, 여기에 배경음악, 워터마크 효과, 그레이스케일 효과가 트랙에 전역적으로 추가된 예이다. 실제 동영상을 만들기 위해서는 섬세한 디자인 적용을 위해 각종 파라미터들이 추가되어 훨씬 복잡하지만, 기

```

<mlt>
  <producer id="producer0">
    <property name="resource">./res/01.jpg</property>
  </producer>
  <producer id="producer1">
    <property name="resource">./res/02.jpg</property>
  </producer>
  <playlist id="playlist0">
    <entry producer="producer0" in="0" out="49" />
    <blank length="50" />
  </playlist>
  <playlist id="playlist1">
    <blank length="50" />
    <entry producer="producer1" in="0" out="49" />
  </playlist>
  <tractor id="tractor0">
    <multitrack>
      <track producer="playlist0" />
      <track producer="playlist1" />
    </multitrack>
  </tractor>
  <producer in="0" out="99" id="audio_bgm">
    <property name="resource">./audio/in_tune.mp3</property>
  </producer>
  <playlist id="playlist_audio">
    <entry producer="audio_bgm" />
  </playlist>
  <tractor id="tractor1">
    <multitrack>
      <track hide="audio" producer="tractor0" />
      <track hide="video" producer="audio_bgm" />
    </multitrack>
    <filter>
      <property name="mlt_service">watermark</property>
      <property name="resource">./watermark1.png</property>
    </filter>
    <filter>
      <property name="track">0</property>
      <property name="mlt_service">grayscale</property>
    </filter>
  </tractor>
</mlt>
    
```

그림 9. 합성된 MLT XML 파일의 예
 Fig. 9. Example of composed MLT XML data

본 구성은 그림 9와 동일하여 복잡한 MLT구조에서도 파싱 및 요소 값의 변경을 통해 장면요소 간 조합 및 필터추가가 가능하다.

그림 10은 사용자정보, 사용자의도 정보 등이 비정형메타태그 형태로 입력되었을 때 이를 단어 단위로 구분한 예이다. 구분된 단어들은 하나의 단어 집합으로 구성하여 스토리보드의 장면과 장면요소들을 연결하는데 사용한다. 그림 10의 예시에서는 “Cafe, Coffee, Americano, LowPrice, Cafe, Icecream, Dessert, Seoul, BundangGu, Silent, Warm, Sweet, Location, Menu, Price, Happy, Daily, Cost, Night, MidNight, Xmas, Christmas”라는 단어가 단어 집합으로 구성되었으며, 이를 기 확보된 장면템플릿의 집합에 질의를 하면 그림 11과 같이 유사도에 따라 결과가 출력됨을 확인할 수 있다. 유사도(Similarity) 값은 [8]의 주제모델링 방식

Context Information

General Context	Cafe
Strategy	Coffee, Americano, LowPrice, Cafe, Icecream, Dessert
Location	Seoul, BundangGu
Mood	Silent, Warm, Sweet
Intention	Location, Menu, Price, Happy, Daily, Cost
Time slot for play	Night, MidNight, xmas, Christmas

[Write/Update] [Cancel]

그림 10. 전략메타정보의 예
Fig. 10. Example of the Strategic Meta Information

의 분류 및 추천 방법을 사용함에 따라 식 (6)과 같이 코사인 유사 방법으로 계산된다. 추천된 장면템플릿을 살펴보

Recommended Scene Templates by Topic Modeling & Recommendation

TID=00291



[Strategic] Cafe
Similarity = [0.728332]

Cafe scene template (Cafe, Coffee, Americano, Icecream, Dessert)

00:30 [Icons]

TID=00160



Dessert Cafe
Similarity = [0.696988]

Cafe scene template (Americano, Coffee, Cafe, Icecream, Dessert)

00:00 [Icons]

TID=00290

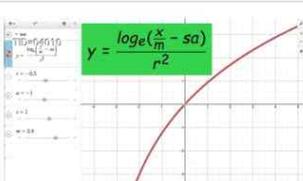


Cafe
Similarity = [0.63747]

Cafe scene template (Cafe, Coffee, Americano)

00:30 [Icons]

TID=00110



Xmas
Similarity = [0.613582]

Xmas template (Xmas, Christmas)

00:30 [Icons]

TID=00070



Menu Scene
Similarity = [0.396984]

Menu scene template (Menu, LowPrice, Cost, Cheap, Night, menu)

00:45 [Icons]

TID=00006



Happy Daily
Similarity = [0.359843]

Daily (children, Warm, Daily, Silent, Happy)

00:45 [Icons]

그림 11. 주제모델링 방법[8]을 사용하여 추천된 장면템플릿 예
Fig. 11. Example of Recommended Scene Templates



그림 12. 장면요소들의 조합방법으로 생성한 출력 비디오 예시
 Fig. 12. Example of Output Video using Scene Element Composition

면 “Cafe, Dessert, Location, Menu, Xmas”와 같은 주제와 부합함을 확인할 수 있다. 이런 과정을 통해 최종 합성된 영상의 예시는 그림 12에 나타내었다.

제안방법을 서비스 제공의 편의성과 결과물의 다양성 측면에서 분석하면 기존 방식으로는 전문디자이너가 30초 분량의 장면 장면템플릿 1건을 새롭게 만들기 위해 평균 30시간 정도의 디자인 시간을 요구했다. 또한 이를 주제에 따라 수정하는 작업도 4~8시간을 필요로 했다. 그러나 제안방법을 이용하는 경우 초기 장면템플릿들을 디자인하는데 소요되는 시간은 비슷했지만, 이를 스토리보드에 따라 자동으로 재조합하여 사용할 수 있기 때문에 결과물의 재생 시간을 쉽게 조절할 수 있었다. 결과물의 다양성 확보를 위해 본 논문에서는 30초 분량의 장면템플릿 30개를 제작했으며, 이를 활용하여 5~10초 분량의 조합용 템플릿들을 “Opening”, “Main Body”, “Ending” 으로 구분하여 각각 10개씩을 제작했다. 3가지로 구분된 장면템플릿들을 무작위 조합하여 영상 합성을 수행하면 이론상으로는 1,000개의 다른 결과물이 만들어 질 수 있고, 길이를 늘이면 조합의 수는 훨씬 늘어나게 된다. 하지만 템플릿 자체가 특정 업종에 특화된 경우에는 조합된 결과가 원래의 목적에 부합하지 못하는 경우도 발생한다. 예를 들어 미용실 소개용 장면 템플릿이 음식점 관련 영상에 사용되는 경우에는 주관적으로 느껴지는 품질과 만족도가 크게 감소했다. 이런 부수효과를 고려하여 조합의 다양성은 유사도 계산 값을 기반으로 사용자가 결정하고 상위 유사도 값을 갖는 장면템플릿을 조합에 사용할 수 있도록 하였다.

본 논문에서 제안하는 방식은 장면요소들을 상호 조합하고 전환효과, 필터효과, 배경음악을 추가할 수 있도록 함을

특징으로 한다. 이 기능은 하나로 구성된 완성형 템플릿뿐만 아니라 본 논문에서 제안한 스토리보드에 따라 조합된 콘텐츠에 대해서도 적용할 수 있다. 이러한 특징적 요소가 전체 성능에 미치는 영향을 확인하기 위해 기술 개발에 직접 참여하지 않은 9명의 피실험자에 대해서 설문 방식으로 만족도 조사를 수행했다. 질문은 크게 2가지로써, 제안방법에 의해 동영상의 품질이 더 좋아졌는지를 질문하는 문항에 {“더 나빠졌다”(0점), “거의 비슷하다”(5점), “더 좋아졌다”(10점)}는 답변 중에서 선택하게 하고, 목표로 하는 콘텐츠에 대한 스토리와 목적 강조 효과를 확인하기 위해 어떤 목적을 가지고 만들어진 것 같느냐는 질문에 대해 {“모르겠다”(0점), “약간 느껴진다”(5점), “확실히 느껴진다”(10점)}라는 답변 중에서 선택하도록 했다. 실험에 사용한 콘텐츠는 그림 11에서 카페 영상을 대조군 영상으로 하여, 그림 12와 같이 크리스마스 시즌을 감안하여 눈이 내리는 효과 및 워터마크, 크리스마스 캐럴 배경음악을 추가한 실험군 영상을 사용했다. 실험결과 콘텐츠 만족도는 평균 2.8점으로써 원본 영상 보다 나빠졌다는 의견이 많았다. 특히 자동으로 부가한 워터마크가 당초의 디자인 품질을 저해하고 이를 해결하기 위해 전문 디자이너가 만든 콘텐츠 요소가 더해진다면 품질 향상도 가능하겠다는 의견이 있었다. 목적 강조 효과와 관련해서는 응답 평균이 5.0점을 보임으로써 다수의 피실험자가 실험군 영상에 대해서 의도와 목적을 느낄 수 있다고 답했다. 제한된 실험 환경에서의 설문이었지만 콘텐츠 디자인의 중요성을 재확인 했다. 만족도 분석은 보다 다양한 콘텐츠와 다수의 사용자 피드백 정보가 필요하고 이를 객관적으로 분석해야 하는 도전적인 문제이다. 현재 서비스 품질 개선 및 안정성 향상을 위한 작업

이 진행 중이므로 향후 더욱 정확한 만족도 평가가 가능해질 것으로 기대한다.

V. 결 론

본 논문에서는 제한된 개수의 장면템플릿들과 각종 장면 요소들을 이용하여 새로운 스타일의 동영상을 생성하는 방법을 제안했다. 동영상을 제작하려는 사용자가 사진이나 동영상을 온라인 시스템 상에 업로드 하면 시스템은 스토리보드에 따라 장면요소에 해당하는 장면템플릿들을 선택 및 배치하고, 내부 리소스를 수정하며, 전략에 맞는 필터효과나 배경음악을 변경하도록 하는 방식이다. 제안 방법은 마케팅 및 홍보전략 의도 등을 이용하여 홍보동영상을 쉽게 제작하는 온라인 서비스 시스템에 적용하여 실제 동영상 합성을 확인했다. 합성결과에 대한 만족도 분석은 향후 수행해야 할 도전적인 문제로서 현재 서비스 품질 개선 및 안정성 향상을 위한 작업이 진행 중이므로 향후 더욱 정확한 만족도 평가가 가능해질 것으로 기대한다.

참 고 문 헌 (References)

[1] E. Jung, J. Choi, Y. Shon, "Influence of advertising type and gender on advertising effectiveness: The efficacy of storytelling advertising", Korean Journal of the science of Emotion & sensibility, pp.489-502,

2012
 [2] Michael Miller, YouTube for business: Online video marketing for any business, 2nd Edition, Pearson Education, 2011
 [3] David Meerman Scott, The new rules of marketing and PR: how to use social media, blogs, news releases, online video, and viral marketing to reach buyers directly, John Wiley & Sons, 2009
 [4] Beaver, Frank Eugene, Dictionary of Film Terms: The Aesthetic Companion to Film Art., Peter Lang Publishing. pp. 46. ISBN 978-0-8204-7298-0, 2006
 [5] Nemo Nemo TV, <http://nemonemo.tv/web/main.php>
 [6] J. Park, M. Park, K. Kim, J. Jeong, J. Lee, and T. Lim, "Development of Cloud-based Service Framework for Promotional Video Synthesis", KICS, 2015 Winter Conference, pp. 795-796, 2015
 [7] J. Park, K. Kim, J. Jeong, and T. Lim, "Development of Processing Interface for Promotional Video Making System based on MLT Multimedia Framework", 2015 Korean Society of Broadcast Engineers Summer Conference, w2.4-4, pp. 1-2, 2015
 [8] J. Park, H. Lee, K. Kim, J. Jung, and T. Lim, "Classification and Recommendation of Scene Templates for PR Video Making Service based on Strategic Meta Information", Journal of broadcast engineering(JBE), Vol 20, No. 6, pp. 848-861, 2015
 [9] Online Standard Korean Dictionary, <http://stdweb2.korean.go.kr/main.jsp>
 [10] [https://en.wikipedia.org/wiki/Plot_\(narrative\)](https://en.wikipedia.org/wiki/Plot_(narrative))
 [11] <https://ko.wikipedia.org/wiki/스토리보드>
 [12] <https://en.wikipedia.org/wiki/Storyboard>
 [13] MLT Multimedia Framework, <http://www.mltframework.org/>
 [14] Deerwester, Scott C., et al. "Indexing by latent semantic analysis." JASIS 41.6 pp.391-407, 1990
 [15] G. Salton and M. McGill, editors. Introduction to Modern Information Retrieval. McGraw-Hill, 1983
 [16] Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." the Journal of machine Learning research 3, pp. 993-1022, 2003

저 자 소 개



박 종 빈

- 2004년 : 성균관대학교 정보통신공학부 학사
- 2006년 : 성균관대학교 전자전기컴퓨터공학부 석사
- 2008년 : 캐나다통신연구소(CRC) 방문연구원
- 2011년 : 성균관대학교 전자전기컴퓨터공학부 박사
- 2012년 : 성균관대학교 정보통신공학부 박사후 연구원
- 2012년 ~ 현재 : 전자부품연구원 스마트미디어연구센터
- ORCID : <http://orcid.org/0000-0002-6123-5300>
- 주관심분야 : 멀티미디어신호처리, 데이터 압축, 정보추론, 머신비전, 정밀제어 시스템 기술

저 자 소 개



김 경 원

- 2001년 : 한국외국어대학교 컴퓨터공학과 학사
- 2003년 : 한국외국어대학교 컴퓨터공학과 석사
- 2013년 : 건국대학교 컴퓨터·정보통신공학과 박사수료
- 2004년 ~ 현재 : 전자부품연구원 스마트미디어연구센터
- ORCID : <http://orcid.org/0000-0001-6530-8426>
- 주관심분야 : 메타데이터, 스마트 TV, N-스크린 서비스, 맞춤형방송, 멀티미디어 검색



정 종 진

- 1997년 : 성균관대학교 정보통신공학부 학사
- 2002년 : 성균관대학교 전자전기컴퓨터공학부 석사
- 2002년 ~ 현재 : 전자부품연구원 스마트미디어연구센터
- ORCID : <http://orcid.org/0000-0003-3924-8938>
- 주관심분야 : 멀티미디어신호처리, 빅데이터 기반 스마트 서비스, 클라우드 기반 스마트홈 플랫폼



임 태 범

- 1995년 : 서강대학교 물리학과 학사
- 1997년 : 서강대학교 전자계산학과 석사
- 2002년 : 대우전자 영상연구소 전임연구원
- 2012년 : 건국대학교 컴퓨터공학과 박사
- 2002년 ~ 현재 : 전자부품연구원 스마트미디어연구센터장
- ORCID : <http://orcid.org/0000-0003-1173-6606>
- 주관심분야 : 홈네트워크 솔루션, IoT, 클라우드 방송, DTV방송, 맞춤형방송, 멀티미디어 검색, IPTV