JDCS
디지털콘텐츠학회논문지
Journal of Digital Contents Society
Vol. 18, No. 7, pp. 1249-1257, Nov. 2017

가상현실 기반 인지재활 콘텐츠를 위한 영상 인식 및 군집화

최 권 택
강남대학교 소프트웨어응용학부 가상현실전공

Image Recognition and Clustering for Virtual Reality based on Cognitive Rehabilitation Contents

KwonTaeg Choi
Division of Software Application, Kangnam University, YongIn-si 16979, Korea

[요 약]

4차 산업혁명과 초고령 시대로 인해, 가상현실을 의료 분야에 적용하려는 많은 연구가 진행되고 있다. 특히 치매에 대한 논의가 활발히 이루어지고 있다. 본 논문은 치매로 인해 발생하는 인지 및 신체 장애를 개선하기 위해 영상 인식 및 군집화 방법을 사용한 가상현실 인지재활 콘텐츠를 제안한다. 기존 인지재활 시스템과는 달리 본 논문에서는 치료 대상자의 추억이 반영되어 있는 여행사진을 사용한다. 자동화된 인지재활 콘텐츠 생성을 위해 사진으로부터 인물정보, 음식사진 여부, 장소 정보, 시간 정보를 추출하고, 군집화를 위해 정규화가 수행된다. 그리고 가상현실 공간에서 여행 사진을 활용해 인지재활 및 신체재활을 강화할 수 있는 시나리오를 제시한다.

[Abstract]

Due to the 4th industrial revolution and an aged society, many studies are being conducted to apply virtual reality to medical field. Research on dementia is especially active. This paper proposes virtual reality based on cognitive rehabilitation contents using image recognition and clustering method to improve cognitive and physical disabilities caused by dementia. Unlike the existing cognitive rehabilitation system, this paper uses travel photos that reflect the memories of the subjects to be treated. In order to generate automated cognitive rehabilitation contents, we extract face information, food pictures, place information, and time information from photographs, and normalization is performed for clustering. And we present scenarios that can be used as cognitive rehabilitation contents using travel photos in virtual reality space.

색인어 : 인지재활, 딥러닝, 영상 인식, 사진 군집화, 가상현실

Key word : Cognitive Rehabilitation, Deep Learning, Image Recognition, Photo Clustering, Virtual Reality

http://dx.doi.org/10.9728/dcs.2017.18.7.1249

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-CommercialLicense(http://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 31 October 2017; **Revised** 21 November 2017
Accepted 25 November 2017

***Corresponding Author; KwonTaeg Choi**

Tel: +82-031-280-3660
E-mail: kwontaeg.choi@kangnam.ac.kr\

Ⅰ. 서론

가상현실은 기술적 가능성을 넘어 엔터테인먼트 분야 외에도 국방, 건강, 교육, 쇼핑, SNS, 의료등 다양한 분야에 응용되고 있다[1][2][3]. 특히 여러 응용 중 의료분야에 가상현실을 적용하려는 다양한 연구가 최근 학계와 병원 등 전문기관에서 이루어지고 있다. 의료에 적용되는 가상현실은 크게 영상 진단학, 외과학, 정신신경과학, 재활의학 분야로 구분될 수 있다

최근 사회가 고령화 시대에 접어들면서 가상현실 기술을 인지재활 치료에 적용하기 위한 연구가 진행되고 있다[4][5][6][7]. 국내의 경우 2026년에 65세 이상 노인 비율이 20%에 달하는 초고령 사회가 예상되고, 치매국가책임제에 대한 논의가 이루어지고 있기 때문에 치매로 인해 발생하는 인지 및 신체장애를 개선하기 위한 관련 연구가 매우 필요한 시점이다.

가상현실 기반의 인지재활 치료 시스템은 가상의 공간을 연출할 수 있고, 정교한 3D 인터페이스를 통해 가상의 물체와 상호작용이 가능해 복잡한 신체 움직임을 표현할 수 있다. 더구나 가상현실은 몰입감이 뛰어나기 때문에 재활치료에 중요한 요소인 자발적인 참여를 이끌어 낼 수 있다.

기존 컴퓨터를 이용한 인지재활 시스템의 경우 매우 오랫동안 연구되었고[8][9], 현재 전문 기관에서 임상적으로도 사용되고 있다. 자발적인 반복학습, 수행결과에 대한 즉각적인 피드백, 객관적인 수행평가 관리가 가능하기 때문에 국내 인지재활 프로그램을 운영하는 기관 중 70% 이상이 컴퓨터 기반 인지재활 시스템을 사용하고 있다[10][11]. 따라서 컴퓨터를 이용한 인지재활 시스템을 가상현실과 결합할 경우 인지재활 치료의 효율성을 높일 수 있을 것으로 기대된다.

그러나 현재까지 연구된 인지재활 시스템에서 사용되는 콘텐츠는 대부분 객관적인 사실로 구성되어 있다. 예를 들면 그림 1의 한국형 전산화 인지재활 프로그램 CoTras[8]의 경우 간단한 영상이 주어지면 사용자는 키보드 형태의 버튼을 눌러서 주어진 문제를 해결해야 한다.

그림 1. 한국형 전산화 인지재활 프로그램 CoTras
Fig. 1. Korean Computer-Based Cognitive Rehabilitation Program(CoTras)

인지재활 프로그램에서 사용하는 콘텐츠는 여러 유형으로 분리된다. 가장 기본적인 유형은 그림 1처럼 사진을 보여주고, 사진과 관련된 단어를 선택한다. 혹은 반대로 용도를 설명해 주고, 용도에 맞는 사진을 선택하기도 한다. 기억력 훈련을 위해서 다수의 사진을 보여주고, 시간이 지나서 동일 사진을 다시 선택하는 유형도 있다. 조금 더 복잡한 인지 능력을 검사하기 위해서 다수의 단어와 이와 관련된 사진을 순서대로 배치하는 유형도 있다. 이는 기억력, 집중력, 순발력을 모두 필요로 한다. 대상자는 이러한 상황에서 개별 대상을 인지해 이들 간의 관계를 파악해야 하는 종합적인 인지 능력이 필요하다. 이렇듯 영상은 인지재활 치료의 기본 콘텐츠로 많이 사용되고 있다.

본 연구에서는 개인의 추억이 반영되어 있는 여행사진을 사용해 회상 기법을 활용한 가상현실 기반 인지재활 시스템을 제안한다. 위에서 살펴본 것처럼 기존 인지재활 시스템에서 사용되는 객관적인 사진이 아닌 가족, 친구 등의 지인과 함께 여행에서 찍은 사진을 콘텐츠로 활용한다. 예를 들어 다수의 과일을 보여주고, 특정 과일을 선택하는 문제대신 지인과 모르는 사람이 있는 사진이 주어지고, 지인만 선택하는 인지재활 문제가 제시될 수 있다. 인지재활 훈련외에도 가상현실 장비를 착용하고 과거의 사진을 감상하는 것은 모니터 화면으로 사진을 감상하는 것과는 몰입감에서 차이가 있다.

이렇게 사진을 이용한 방법은 사진치료 분야에서 이미 사용되고 있다[12]. 사진의 매체적 속성(기록성, 사실성, 전달성)을 활용한 치료기법은 신뢰감 형성, 기억 재생, 인지방식에 도움이 되기 때문에 사진 치료 기법은 소외되는 노인들의 자존감과 삶의 질 향상이 가능한 것으로 알려져 있다. 특히 치매 노인 회상 요법에 관한 연구는 오랫동안 이루어져 왔고, 이는 노인의 우울감을 감소시키는 효과와 더불어 회상 기능의 일부 기능을 향상시키는데 효과적이라고 알려져 있다. 그러나 치매 노인의 개인별 특성 및 수준에 적합한 맞춤형 단계별 회상을 이용한 인지 훈련 프로그램 개발과 그에 따른 임상적 적용에 대한 연구는 매우 미비한 상황이다.

본 논문에서는 사진의 내용을 자동으로 인식해 몰입형 가상현실 기기에서 인지재활 치료를 할 수 있는 시스템을 제안한다. 이를 통해 사진이 가지고 있는 회상 효과를 활용하고, 가상현실 공간에서 이를 구현함으로써 치료에 지속적으로 몰입할 수 있도록 하고자 한다.

본 논문의 구조 다음과 같다. II 절에서는 제안하는 시스템의 전체 흐름도, 사진 분석 및 군집화 알고리즘 , 그리고 이를 활용한 인지재활 콘텐츠 사례를 설명한다. III 절에서는 실험을 통해 영상 분석의 정확도와 지역정보 군집화에 대한 실험을 수행한다. 마지막으로 결론은 IV 절에서 도출한다.

Ⅱ. 제안방법

2-1 제안하는 시스템 개요

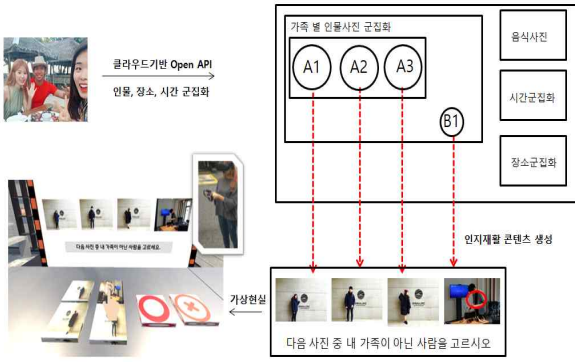

그림 2. 제안하는 인지재활 시스템 개요
Fig. 2. Outline of the proposed cognitive rehabilitation system

제안하는 시스템은 그림 2와 같다. 인지재활 콘텐츠로 사용될 사진이 가족 구성원에 의해 주어지면 사진속의 메타 데이터(시간, 장소) 및 문맥 데이터(인물, 나이, 성별, 음식여부)를 처리하는 클라우드 기반 OpenAPI를 사용해 사진들을 분류하고 군집화를 수행한다. 이러한 정보를 사용해 인지재활의 목적에 맞는 콘텐츠를 생성하고 가상현실 공간에서 활용하게 된다.

그림 2의 상황은 내가 알고 있는 사람과 모르는 사람을 정확히 인지해 구분해야 하는 시나리오이다. 4 장의 사진이 문제로 주어지는데 3장의 사진은 내 가족사진으로 군집화된 영역(A1, A2, A3)에서 선택하고, 나머지 한 장은 다른 가족의 군집화된 영역(B1) 으로부터 선택해 인지재활 문제로 만들어진다. 이때 사용자가 선택해야 하는 답은 B1에서 선택된 사진이다. 이렇게 문제가 만들어지면 인지재활 치료 대상자는 가상현실 공간에서 주어진 문제를 해결해야 하는데 이때 단순 선택 방식 보다는 가상현실 공간상에 배치하여 신체활동을 충분히 할 수 있도록 유도한다. 따라서 치료 대상자는 문제를 풀어가면서 사진을 회상하고, 인지적 활동을 하게 되고 동시에 신체를 움직이게 된다.

그림 2의 인지재활 콘텐츠를 만드는 과정에서 얼굴 사진이 같은 사람끼리 군집화 되어 있지 않으면 동일한 사람의 사진이 선택되어 혼란을 초래할 수 있고, 문제 난이도 조절에 어려움이 발생한다. 따라서 최대한 얼굴의 유사도가 다른 사진 중에서 3장의 사진이 선택되어야 한다. 따라서 제안하는 시스템에서는 얼굴 사진에 대한 군집화가 되어야 한다. 이는 다른 유형에 대해서도 마찬가지이다. 다음절에서는 사진으로부터 인물, 음식, 시간, 장소에 대한 군집화 방법에 대해서 설명한다.

2-2 인물 및 음식 사진 군집화

일반적으로 사람들은 스마트폰에 많은 양의 사진을 보관하고 있다. 취미로 촬영하는 사람도 있고 특별한 이벤트가 아닌 일상생활 전반에 걸쳐 사진을 촬영하기도 한다. 본 논문에서 인지재활치료에 사용할 사진은 전문가 촬영이 아닌 스마트폰으로 촬영하는 사진을 분석대상으로 한다. 스마트폰으로 촬영한 사진들은 인물사진(셀피포함), 음식사진, 풍경사진, 기념사진(장소나 건물)이 대부분이다. 본 논문에서는 인물, 음식, 장소 정보를 활용했다. 각각의 사진에 따라 인지재활 콘텐츠가 달라지기 때문에 이들 사진을 정확하게 구분할 필요가 있다.

- 인물 사진 분류 및 군집화

본 논문에서는 얼굴이 검출되는 경우 인물사진으로 분류했다. 그리고 나이와 성별정보를 사용해 다양한 유형의 시나리오를 생성할 수 있기 때문에 검출된 얼굴에 대해 나이와 성별 정보를 추가적으로 인식했다. 예를 들면 여러 아이 사진 중 가족과 관련이 없는 다른 아이를 구분해내는 문제, 남자 두명과 여자 1명이 있는 사진을 빠르게 기억해서 선택해야 하는 문제등 다양한 시나리오를 생각해 볼 수 있다.

정확한 얼굴 군집화를 위해서는 얼굴 인식 기술을 사용하면 된다[13][14]. 그러나 이는 교사학습방법으로 사용자 혹은 가족들이 많은 학습데이타를 사람별로 구분해서 제공해야 하기 때문에 인지재활 시스템에서 사용하기는 쉽지 않다. 따라서 본 논문에서는 비교사 학습기반으로 얼굴을 군집화하기 위해 성별과 나이 정보를 사용했다.

사진으로부터 성별과 나이를 추정하기 위해서 Naver Clova Face Recognition API를 사용했다. 성별의 경우 신뢰도에 대한 확률값이 함께 주어진다. 나이도 특정 나이가 아닌 30-33형태의 범위데이터와 역시 신뢰도가 함께 주어진다.

본 논문에서는 가족 사진내에서 인물 사진을 군집화하기 성별과 나이에 대해 K-Means 알고리즘을 사용한다. 이를 위해 성별에 대한 정보는 식 (1) 처럼 변환하였다.

$$f_{(gender)} = \begin{matrix} 0 & \text{if } c > \Theta \text{ and } male, \\ 1 & \text{if } c > \Theta \text{ and } female, \\ 2 & \text{if } c > \Theta \text{ and } child \end{matrix} \qquad (1)$$

여기서 c는 Naver API에서 성별 추정에 대한 신뢰도 값이고, Θ는 최소 신뢰도를 위한 역치(Threshold)이다. 이는 실험적으로 정해진다. 어린이의 경우 성별추정이 어렵기 때문에 Naver API는 성별결과로 child를 반환한다.

나이의 경우는 범위로 주어지기 때문에 식 (2)처럼 최소, 최대값의 평균으로 계산한다.

$$f_{(age)} = \frac{(\min + \max)}{2} \quad \text{if } c > \eta \qquad (2)$$

여기서 c는 Naver API에서 나이 추정에 대한 신뢰도 값이고, η는 최소 신뢰도를 위한 역치이다. 이는 실험적으로 정해진다. 한 가족 내에서 인물 군집화를 위해 식 (1)과 식 (2) 두 개의 특징을 사용해 K-Means 알고리즘을 사용했다.

이렇게 인물사진을 구분하고, 성별과 아이를 구분하고 군집화가 되면 다양한 인지재활 콘텐츠를 생성할 수 있다. 그림 3은

그림 3. 인물 사진 분석 및 군집화를 이용한 시나리오 예
Fig. 3. Scenario example using face analysis and clustering

이들 정보를 복합적으로 활용한 시나리오이다. 치료 대상자는 제공되는 사진을 감상하면서 모든 사진을 4개로(아이 사진, 남자 사진, 여자 사진, 그리고 내가 모르는 사람) 구분해야 한다. 치료 대상자는 이렇게 기억속의 사진을 감상하고, 인지한 후 몸을 움직여 인지재활 치료과정을 수행한다.

- 음식 사진 분류

여행은 먹는 즐거움이라 할 만큼 한국 사람들은 여행지에 대한 경험과 더불어 맛있는 음식을 먹고 이를 촬영해 사람들과 공유한다. 따라서 음식 사진으로부터 과거의 좋은 경험을 상기할 수 있기 때문에 음식 사진은 인지재활 치료에 좋은 콘텐츠이다.

그러나 현재 주어진 사진이 음식인지 아닌지 구분해야 하는데, 음식만을 전문적으로 인식하는 신뢰성 있는 Open API가 존재하지 않기 때문에 본 논문에서는 Google Vision API 중 LABEL 서비스 기능을 사용해 영상을 단어로 인식하고, 분석된 단어가 음식에 관련된 단어로 이루어져 있으면 음식으로 판단하는 방법을 사용했다. 이를 위해서 음식단어 사전을 구성해야 하는데, 이를 위해 다양한 음식 사진을 모아서 Google Vision API를 사용해 빈도수가 높은 음식 단어들로 음식 단어 사전을 구성했다. 따라서 음식 사진 여부는 식 (3)처럼 결정된다.

$$f_{(food)} = \begin{matrix} 1 \text{ if } label(x) \in food\ dictionalry, \\ 0 \qquad\quad otherwise \end{matrix} \qquad (3)$$

음식 사진을 구분한 후 인지재활 콘텐츠로 활용할 수 있는 시나리오는 다양하다. 음식만 있는 사진의 경우 음식 사진을 모아 여행에서 먹은 음식과 먹지 않은 음식을 구분해야 하는 문제도 가능하고 음식 사진과 관련 없는 인물, 시간, 장소에 관한 연관 단어를 제거하는 문제도 가능하다. 그림 4의

그림 4. 음식 사진 인식을 이용한 시나리오 예
Fig. 4. Scenario example using food image

시나리오는 얼굴이 포함되어 있는 않은 음식 사진과 관련 있는 사진은 음식 사진위에 올려놓고, 관련 없는 사진은 상자에 담아야 하는 유형의 콘텐츠이다. 사진에는 음식을 촬영했던 장소의 이름에 해당하는 카드 1개 , 같은 날짜로 군집화된 영상 중 인물 사진으로 만든 카드 1개, 사진의 계절에 해당 하는 단어 카드 1개, 정답이 아닌 다른 년도에서 선택된 사진 카드 1개로 구성되어 있다. 재활치료 대상자는 이 사진을 보며 언제, 누구와 어디서 음식을 먹었는지 회상해서 문제를 해결해야 한다.

2-3 역지오코딩을 이용한 장소 군집화

스마트폰으로 사진을 촬영할 때의 장점은 촬영 장소의 위치를 알 수 있다는 점이다. 몇 년 전만 하더라도 스마트폰의 GPS 정확도 및 배터리 소모 문제로 사진 촬영 시 GPS 정보를 사진에 추가하는 기능이 활성화하지 않았지만 최근에는 이 기능이 기본적으로 활성화되어 많은 영상에 위치정보가 포함되어 있다. 사진 촬영 후 촬영 지점의 위도, 경도 정보는 사진의 EXIF 부분에 기록된다. 그리고 실내에서 촬영해 GPS 정보를 위성으로 부터 받을 수 없는 경우 가장 최근 위치, 와이파이, 기지국 정보를 통해 위치 파악이 가능하다.

그러나 위도, 경도 정보로 사진을 구글 지도위에 오버레이해서 어디에서 촬영했는지 보여줄 수 있지만 인지재활 치료에서는 위도, 경도 정보가 아닌 지역 이름이 필요하다. 예를 들어 같은 장소에서 찍은 사진끼리 분류하는 문제의 경우 위도, 경도 정보로는 같은 장소인지 알 수가 없다. 장소에 대한 이름이 필요하고 이들이 군집화 되어 있어야 한다. 다른 예로는 국내 여행과 국외 여행을 구분하는 문제도 가능하다. 이때는 국가별 군집화가 필요하다.

일반적으로 도시 이름 이나 지역 이름으로부터 위도 및 경도 정보를 추출하는 것을 지오코딩(Geocoding)이라 하고, 거꾸로

그림 5. 구글 역지코딩 API의 JSON 출력
Fig. 5. JSON output of the Google Reverse Geocoding API

위도, 경도로 부터 지역 이름을 추출하는 것을 역지오코딩(Reverse Geocoding)이라 한다. 지오코딩과 역지오코딩을 위해서는 위도, 경도 정보와 지역 정보에 대한 정확하고 방대한 데이터가 필요한데 구글은 클라우드 서비스를 통해 지오코딩 API와 역지오코딩 API를 모두 지원한다. 그림 5는 특정 지역에 대한 역지오코딩의 JSON(JavaScript Object Notation) 출력 결과 이다.

구글 역지오코딩 API의 경우 나라마다 주소 체계가 복잡하고, 나라마다 다르기 때문에 가능한 모든 경우를 표현할 수 있도록 API가 설계되어 있다. 그림 5를 보면 전체 주소 정보는 formatted_address 속성에 들어가 있고, 시, 도, 구, 군등 다양한 행정 지역 이름을 지원하기 위해 address_components 속성은 배열로 정의되어 있다. 그림 5에서 고양시는 3번째 행정 지역 정보를 뜻한다. types 속성으로 시, 도, 군 등에 대한 정보를 알 수 있고, 실제 지명 이름은 long_name과 short_name 속성을 통해 알 수 있다.

다만 구글 역지오코딩 API결과에서 formatted_address를 지역 이름으로 사용하는 것은 제안하는 시스템에서는 적합하지 않다. 이러한 형태의 정보를 가지고 장소 군집화는 불가능하다. 일반적으로 여행한 지역의 정확한 주소 위치를 기억할 수도 없고, 응용 분야에 따라 정확한 주소가 필요하지 않다. 본 논문에서는 구글 역지오코딩의 JSON 포맷 결과를 분석해 국가 정보와 국내의 경우 동 혹은 건물 이름을 장소에 대한 대표 이름으로 사용하고, 해외의 경우 행정 지역 정보를 대표 이름으로 사용해 장소에 대한 군집화를 수행했다.

JSON에서 x라는 types의 속성 값을 갖는 long_name 속성 값을 반환하는 함수를 식 (4)로 정의하면,

$$h(x) = find(reversegeo(lat, long), types = x) \quad (4)$$

그림 6. 지역 정보 군집화를 이용한 시나리오 예
Fig. 6. Scenario example using location clustering

국가이름은 식 (5)를 통해 알 수 있다.

$$f_{(country)} = h(country) \quad (5)$$

그리고 지역 군집화를 위한 단어는 식 (6)으로 결정된다.

$$f_{(location)} = \begin{cases} h(\text{동}) \text{ if } korea, \\ translate(h(\text{지역})) otherwise \end{cases} \quad (6)$$

여기서 translate()함수는 영문 이름을 한글 이름으로 번역하는 함수이다.

국가 이름은 country types에서 명확하게 알 수 있지만, 나머지 한국주소의 동에 대한 정보와 외국의 경우 지역 이름이 어떤 address_components와 관련되어 있는지 명확하지 않기 때문에 본 논문에서는 실험을 통해 이를 결정하고자 한다.

지역 정보를 군집화하고 인지재활 콘텐츠로 활용할 수 있는 시나리오는 다양하다. 그림 6은 국내에서 촬영한 사진과 해외에서 촬영한 사진이 섞여 있고, 사진을 국가별로 정리하는 시나리오이다. 사진을 담아야 하는 박스에는 사진이 촬영된 나라 이름과 연도가 회상을 도와주기위해서 표시될 수 있도록 구현했다.

2-4 Time 기반 사진 군집화

사진속의 시간 정보는 인지재활 콘텐츠 관점에서 사진을 시간 순서로 배치한다든지 과거와 최근 사진을 구분한다든지 하는데 사용할 수 있다. 인지재활 대상자는 인지 능력에 문제가 있기 때문에, 사진속의 날짜를 정확하게 기억하기는 어렵다. 이는 일반인에게도 고도의 기억 능력을 필요로 한다. 본 논문에서는 EXIF 영역에서 시간정보 date를 읽고 식 (7), 식 (8) 처럼 년도와 계절에 대해 정규화 하고 시간에 대한 군집화를 수행했다.

$$f_{(time)} = year(date) - \arg\min(year(date)) \quad (7)$$

$$f_{(season)} = \begin{matrix} 0 \text{ if } month(date) \in [3,4,5] \\ 1 \text{ if } month(date) \in [6,7,8] \\ 2 \text{ if } month(date) \in [9,10,11] \\ 3 \text{ if } month(date) \in [12,1,2] \end{matrix} \quad (8)$$

년도의 경우 날짜에서 해당 년도만 추출해 전체 데이터에서 최소 년도와 차이를 구해 정규화 했다. 계절의 경우 해당 월에 따라 0~3으로 정규화 했다.

시간 정보를 군집화하고 인지재활 콘텐츠로 활용할 수 있는 시나리오는 다양하다. 가장 대표적인 시나리오는 사진을 순서대 로 정렬하는 것이다. 본 논문에서는 식 (7)에서 년도 정보만으로 군집화 되지만, 식 (8)의 계절 정보를 이용해 군집화 할 수 있기 때문에 3년 간 사진을 사용한다면 최대 12개의 그룹으로 나누어져 시간 순서대로 정리할 수 있도록 콘텐츠를 만들 수 있다.

표 1에 지금까지 논의한 사진에서 추출하고 인식할 수 있는 정보와 군집화하기 위해 정규화된 특징 값, 그리고 대표적인 활용 콘텐츠에 대해서 정리하였다.

Ⅲ. 실험

본 논문에서는 사진으로부터 얼굴, 성별, 나이, 음식사진, 촬영 지역, 촬영 시간 정보를 인식해 인지재활 콘텐츠를 만든다. 얼굴 검출의 경우 오랫동안 연구되었고, 일반인이 스마트폰으로 촬영하는 가족사진 특성상 정면에 가까운 영상이 많다. 더구나 대다수 얼굴 검출기의 성능이 매우 높기 때문에 얼굴 검출 성능은 측정하지 않았다. 또한 시간에 따른 군집화도 명확한 값이라 실험으로써 의미가 없어 별도 실험을 수행하지 않았다. 따라서 성별인식, 나이인식, 가족사진 내 얼굴 군집화, 음식사진 인식, 역지오코딩 군집화에 관한 실험을 수행하였다. 표 2에 각각의 실험에 사용되는 실험 데이터셋에 대해서 정리하였다.

표 1. 분석 정보 및 활용 콘텐츠 유형 요약
Table. 1. Content type summary

영역	인식, 군집화	활용 콘텐츠 예시
얼굴	얼굴위치 성별, 나이 군집화번호	아는 사람 모르는 사람 구분 조건에 맞는 사진 선택
음식	음식사진여부	음식 사진과 관계없는 사진 선택
지역	국가 대표명	사진을 국가별로 정리 같은 장소에서 찍은 사진끼리 분류
날짜	년도 계절	과거와 최근 사진 구분 사진을 시간 순서대로 정렬

표 2. 사진 분석 성능 검증을 위한 데이터셋 특성
Table. 2. Dataset characteristics for photo analysis performance evaluation

실험	DB	테스트영상
얼굴 군집화	가족영행사진	200장 / 2가족
음식사진 군집화	개인사진	300장
	internet 사진	300장
지역 군집화	국내개인사진	100장
	국외개인사진	500장/5개국

얼굴 사진의 경우 스마트폰으로 촬영된 사진 중 얼굴이 포함된 사진을 두 가족으로부터 100장 씩 200장을 수집했다. 사진 파일의 해상도가 너무 크면 Open API에서 분석할 수 없기 때문에 모든 사진의 크기는 2000x1500수준으로 크기를 줄였다. 음식 사진은 개인이 소장하고 있는 음식사진과 무작위로 인터넷으로 수집한 사진을 사용했다. 해상도가 큰 경우 2000x1500수준으로 영상의 크기를 줄였다. 국내 사진과 해외 사진의 경우 GPS 정보가 명확하게 추출되는 사진 500장을 사용했다.

3-1 성별, 나이, 얼굴 군집화

성별, 나이 그리고 가족 구성원 내 얼굴 군집화 성능을 테스트하기 위해 표 3의 데이터셋을 사용해 측정하였다.

우선 Naver Clova Face Recognition API를 사용해 성별인식에 대한 신뢰도를 평가하였다. 2개의 데이터셋 200장의 영상으로부터 검출된 얼굴에 대한 성별인식에 대한 신뢰도에 대한 분포도를 조사했다. 성별 추정에 대한 신뢰도는 매우 높게 나타났다. 본 논문에서는 식 (1)에서 신뢰도에 대한 역치값으로 0.321로 정했다. 역치가 낮은 얼굴 사진은 성별 인식이나 얼굴 군집화가 필요하지 않는 다른 시나리오에서 사용될 수 있다.

실제 성별 추정에 대한 정확도는 표 4에 정리하였다. 성별 인식에 대한 정확도는 다소 성능 개선이 필요해 보인다.

표 3. 성별 인식, 나이 인식, 가족 구성원 내 얼굴 군집화 성능 검증을 위한 데이터셋 및 특성
Table. 3. Dataset characteristics for gender recognition, age recognition and face clustering performance evaluation

DB	인원	성별특성	나이특성
family #1	4	남1, 여1,아이2	40대, 40대, 10세미만
family #2	7	남3, 여4	50대, 40대, 30대,20대

표 4. 성별 추정에 대한 정확도 평가
Table. 4. Accuracy evaluation of gender estimation

성별	남자	여자	아이	평균
인식률	93.2%	83.2%	86.6%	87.7%

표 5. 나이 추정에 대한 정확도 평가
Table. 5. Accuracy evaluation of age estimation

DB	아이-10대	20대-30대	40대-50대
RMS에러	5.2	10.3	14.87

다음 실험은 나이인식에 대한 신뢰도를 평가하였다. 2 개의 데이터셋 200장의 영상으로부터 분석한 나이 인식의 신뢰도를 분석했다. 나이에 대한 신뢰도 값은 0.001 ~ 0.5445까지 그 범위가 매우 크기 때문에 신뢰도 값을 활용하기는 불가능해 보였다. 따라서 나이 추정을 위해 식(2)에서 역치값으로 0.0을 사용했다.

실제 나이 추정에 대한 정확도는 표 5에 정리하였다. 사진 속 사람의 나이를 정확히 알고 있기 때문에 영상 인식 엔진이 추정한 나이와 실제 나이 차이를 기반으로 RMS(Root Mean Square) 에러를 계산했다. Naver Clova Face Recognition API가 나이 인식의 결과를 범위값으로 (min ~ max)로 제공하기 때문에 실험에서 min, max에 대한 평균(mean)값을 추정 나이로 사용했다. 신뢰도에서 나타나듯 나이 추정에 대한 인식률은 높지 않은 것으로 나타났다.

다음 실험은 성별과 나이를 특징으로 사용해 K-Means 군집화를 수행해 가족 구성원 사이의 얼굴 군집화가 가능한지 실험하였다. K-Means를 수행할 때 가족 구성원에 대한 전체 가족이 몇 명인지는 알고 있다고 가정했다. 그림 7은 family set #2 에 대한 K-Means 군집화 결과이다. x 축은 나이이고, y 축은 성별이다.

그래프의 결과를 보면, 남자, 여자, 아이에 대한 인식 성능이 비교적 높기 때문에 y축으로 잘못 군집화되지는 않았다. 같은 성별에서는 나이차가 크지 않은 경우 구분력이 떨어졌다. 향후 좀 더 정확하게 나이를 추정할 수 있는 정교한 방법이 사용되거나 영상 분석 기반의 얼굴 군집화 알고리즘과 결합해 인식률을 높일 필요가 있다.

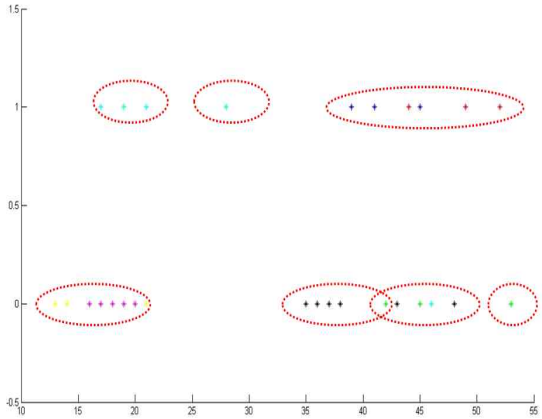

그림 7. 성별과 나이를 이용한 가족사진 얼굴 군집화
Fig. 7. Family photo face clustering using gender and age

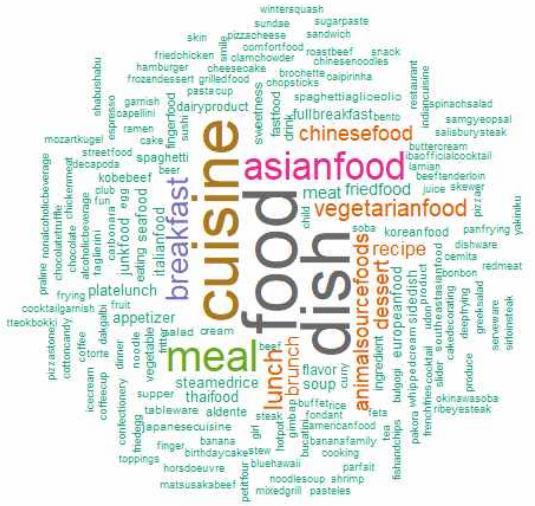

그림 8. 음식 사진에 대해 구글 비전 API 결과에 대한 단어구름
Fig. 8. Word cloud of Google Vision API results in food photos

3-2 음식 사진 분류

음식 사진 여부를 인식하기 위해 음식 사진 300장은 구글의 비전 API를 사용해 상위 10개의 레이블을 요청했다. 레이블에 대한 분포를 그림 8에 표시하였다.

음식 사진에 대해 구글 비전 API가 추정한 단어에서 빈도수가 높은 단어는 food, cuisine, dish이다. 이외에도 launch, breakfast 등 식사 시간에 대한 추정이 많고, meal, fruits. dessert , appetizer 등 음식 종류에 해당하는 경우도 많았다. 그러나 음식 종류는 너무 많기 때문에 음식 사전 데이터로 사용하기는 어렵다. food의 경우도 나라별, 종류별로 구분되기 때문에 여기서는 모두 food로 평가하였다. 음식 관련 단어는 많지만 구글이 사용하는 단어는 한정적이다. 따라서 가장 빈도수가 많은 단어를 선택해서 음식사전을 만들었다. 본 논문에서는 빈도수가 높은 상위 5개의 단어를 포함하는 경우 food로 구분하였다. 음식 분류 인식률은 94.5%였다. 음식사진에 대한 인식률은 매우 높은 것으로 보인다.

모든 음식 사진을 인지재활 콘텐츠로 사용할 필요는 없으나 인식률이 낮으면 많은 사진들이 기억회상 요법에서 사용할 수 없기 때문에 인식률은 어느 정도 유지되어야 한다. 실험 결과는 충분한 인식률로 보인다.

3-3 역지오코딩을 이용한 지역정보 군집화

최신 스마트폰으로 촬영한 경우 GPS 정보가 사진에 포함되어 있지만, 예전 스마트폰이거나, 촬영 앱에 따라 GPS 정보 자체가 들어가 있지 않은 경우, (0,0)으로 기록되어 있는 경우, 원본 사진이 아닌 경우, 해외에서 촬영할 당시 서비스 비활성화로 GPS 정보가 들어가 있지 않은 경우도 있다. 본 논문에서는

표 6. 국가에 따른 구글 역지오코딩 JOSN 결과 분석
Table. 6. Google reverse geocoding JOSN results by country

		short_name	long_name
지역정보	국내	한글	한글
	해외	지역언어	영어
국가이름	국내	한글	한글
	해외	영어	한글

GPS 정보가 정상적으로 들어가 있는 사진에 대해서만 군집화를 수행하고 인지재활 콘텐츠로 사용했다.

EXIF 정보로부터 GPS 정보를 추출한 후 구글의 역지오코딩 API를 사용해 위도, 경도 정보를 지명으로 변경하였다. 이때 표준화된 방법이 존재하지 않고, 국가마다 주소 표기법이 다르기 때문에 구글은 자신의 기준으로 주소체계를 구성했다. 따라서 구글이 JSON으로 보내준 데이터에서 어떤 정보를 기준으로 지역 정보를 군집화 할 것인지 실험을 통해 분석했다.

구글의 역지오코딩 API의 결과는 모든 주소 정보가 영어로 반환된다. 제안하는 시스템에서는 한글로 표현되어야 인지재활 콘텐츠 생성을 자동화할 수 있기 때문에 API 호출시 language=ko 옵션을 사용해 한글 주소가 반환되도록 했다. 5개 국가에 대해서 실험을 수행했다. 이렇게 하더라도 한글, 영어, 자국어가 섞여져 나오기 때문에 주소에 대한 언어 표기법에 관한 실험이 필요했다. 언어 표기법에 대한 실험결과를 표 7에 정리하였다.

표 6의 실험 결과로부터 지역 정보를 이름으로 군집화하기 위해 short_name보다는 long_name이 적합해 보인다. 국내 주소의 경우 모든 주소 정보에 대해 short_name, long_name 모두 한글지명으로 반환된다. 해외의 경우 long_name은 영어로, short_name은 지역 언어로 나오게 된다.

따라서 long_name을 사용해 지명에 대해서 한글 이름을 사용해 군집화하는것이 유리하다. 다만 해외의 경우 국내처럼 장소 기반으로 콘텐츠를 만들기 위해서는 이를 다시 한국어로 번역이 필요하다. 본 논문에서는 영어로 반환되는 지역정보를 한글로 표현하기 위해서 구글의 번역 API를 사용했다. 이를 통해 영어식 표현은 한글로 변환되어 long_name 으로 반환되는 모든 값을 한글 단어로 군집화 했다.

실험을 통해 결정해야 하는 두 번째 요소는 장소를 대표할 수 있는 이름이다. 본 논문에서는 국내 주소의 경우, 시는 범위가 넓어 동 범위로 지역 정보를 군집화하고, 해외의 경우 지리에 익숙하지 않기 때문에 시군에 대응하는 이름을 대표 이름으로 사용하기로 했다. 표 7에 구글의 역지오코딩 결과를 주소체계와 관련지어 정리했다.

실험결과로부터 국내의 경우 sublocality_level_2 값을 장소 군집화를 위한 명칭으로 사용할 수 있음을 확인할 수 있었다. 해외의 경우 locality를 장소명으로 사용하면 된다.

표 7. 구글 역지오 코딩 결과
Table. 7. Google's reverse geocoding result

국내외여부		long_name	군집화단어
국내	국가	country	O
	도	administrative_area_level_1	
	시,군	locality	
	구	sublocality_level_1	
	동,면,읍	sublocality_level_2	O
	리	sublocality_level_3	
	거리	sublocality_level_4	
	건물	establishment	O(건물)
해외	국가	country	
	행정지역	locality	O(번역)
	지역이름	sublocality_level_1	

행정지역에 대한 정보 외에 establishment에는 숫자로 이루어진 지번이 들어가거나 장소나 건물이름이 들어가 있기 때문에 establishment가 숫자로 이루어져 있지 않은 경우 군집화 지역이름으로 sublocality_level_2 보다는 establishment를 사용하는게 적합해 보인다.

Ⅳ. 결론

본 논문에서는 기존 컴퓨터를 활용한 인지재활 콘텐츠와 달리 개인의 경험과 추억이 담긴 여행 사진을 이용한 인지재활 콘텐츠를 생성했다. 구체적으로 사진 인식 및 군집화 방법을 제안하고, 이를 몰입형 가상현실에서 활용할 수 있는 시스템도 함께 제안하였다. 인지재활 콘텐츠를 자동으로 생성하기 위해 사진 영상에서 의미 있는 정보를(얼굴 영역, 성별, 나이, 음식사진, 지역정보, 시간정보) 추출해야 하는데 본 논문에서는 네이버와 구글이 제공하는 Open API를 사용해 특징을 인식하고 이를 정규화를 통해 군집화 할 수 있는 방법을 제안하였다. 스마트폰으로 촬영한 사진을 수집해 성능을 검증하고 그 가능성도 함께 살펴보았다. 가상현실과 인지재활을 융합한 연구는 다양한 기술 개발이 필요해보이고, 무엇보다도 다양한 인지재활 콘텐츠 개발과 이에 대한 임상적 연구가 매우 필요한 시점이다.

감사의 글

본 연구는 2017년도 강남대학교 교내 연구비 지원 사업에 의하여 이루어진 연구로서, 관계부처에 감사드립니다.

참고문헌

[1] Mansik Kim, Jungho Kang and Moon-seong Jun, 2016,

"Market and Technical Trends of VR Technologies," The Korea Contents Association Review, Vol. 14, No. 4, pp. 14~16.

[2] K.-E. Chang, J. Zhang, T.-C. Liu, "Workshop on Virtual, Augmented and Mixed Reality in Education (VAMrE 2017) Summary", 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct), Pages: 293 - 293.

[3] Jiyoung Kang, 2017, "Study on Characteristics of Digital Realism Aspect for HMD based Virtual Reality Films," Journal of Digital Contents Society, Vol. 18, No. 5, pp. 849~858.

[4] C. S. Kim, Y. H. Kwon, 2005, “Therapeutic Virtual Reality Program in Chronic Stroke Patients Recovery of Upper Extremity and Neuronal Reorganization", Journal of Special Education & Rehabilitation Science, 44(1), 87-106.

[5] Flynn, D., Van Schaik P., Blackman, T., and Femcott, C., 2003, “Developing a Virtual Reality-Based Methodology for People with Dementia: A Feasibility Study”, CyberPsychology, 6(6), 591-611.

[6] Larson EB, Feigon M, Gagliardo P3, Dvorkin AY, "Virtual reality and cognitive rehabilitation: a review of current outcome research", NeuroRehabilitation. 2014;34(4):759-72.

[7] M. I. Khan, A. Prado, S. K. Agrawal, "Effects of Virtual Reality Training With Trunk Support Trainer (TruST) on Postural Kinematics Sign In or Purchase", IEEE Robotics and Automation Letters (Volume: 2, Issue: 4, Oct. 2017.

[8] Y. G. Kim, 2011, Development of a Korean Computer-based Cognitive Rehabilitation Program (CoTras) for Patients with Cognitive Disabilities and the Validation of it’s Effects. Pusan National University in South Korea, 34-9.

[9] W. H. Kim, J. H. Chang, 2017, “A system for efficient checkup and management of dementia using a smart-phone”, Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology, 7(1): 577-585.

[10] H. R. Lee, J. Y. Kim, and D. S. Han, 2012, “A Survey on the Cognitive Rehabilitation of Occupational Therapy in Korea", Journal of Korean Society of Occupational Therapy, 20(2): 73-84.

[11] S.-H. Han, E.-J. Jo, D.-h. Noh, K.-Y. Kam, "Effects of Korean Computer-Based Cognitive Rehabilitation Program (CoTras) on Frontal-Executive Functions in Patients with Traumatic Brain Injury", Journal of the Korea Academia-Industrial cooperation Society, Vol. 16, No. 5 pp. 3344-3352, 2015

[12] Lee Gayoung and Kim Sunhee K., "A literature review on the use of photography and digital images in art therapy," The Korean Journal of Arts Studies, No. 13, pp. 219~241.

[13] S. Xia, H. Pan, A.K. Qin, "Face Clustering in Photo Album", ICPR 2014.

[14] C. Otto, B. Klare, A. K. Jain, "An Efficient Approach for Clustering Face Images", IAPR International Conference on Biometrics 2015.

최권택(KwonTaeg Choi)
2001년 : 한림대학교 컴퓨터공학과 (공학학사)
2006년 : 연세대학교 컴퓨터과학과 (공학석사)
2011년 : 연세대학교 컴퓨터과학과 (공학박사-컴퓨터비전, 패턴인식)

2011년～2015년: LG전자
2016년～현 재: 강남대학교 컴퓨터미디어정보공학부 교수
2017년～현 재: 강남대학교 소프트웨어응용학부 교수
※관심분야 : 가상현실, 모바일컴퓨팅, 기계학습, HCI