

관람객 얼굴 표정 및 제스처 인식 기반 인터랙티브 디지털콘텐츠의 문제점 추론 시스템

권도형^o, 유정민^{*}

한국전통문화대학교, 문화유산산업학과^o

한국전통문화대학교, 문화유산산업학과^{*}

e-mail: {ehgud1, jmyu}@nuch.ac.kr^{o*}

Problem Inference System of Interactive Digital Contents Based on Visitor Facial Expression and Gesture Recognition

Do-Hyung Kwon^o, Jeong-Min Yu^{*}

Dept. of Cultural Heritage Industry, Korea National University of Cultural Heritage^o

Dept. of Cultural Heritage Industry, Korea National University of Cultural Heritage^{*}

● 요약 ●

본 논문에서는 관람객 얼굴 표정 및 제스처 인식을 기반으로 인터랙티브 디지털콘텐츠의 문제점 추론 시스템을 제안한다. 관람객이 콘텐츠를 체험하고 다른 장소로 이동하기 전까지의 행동 패턴을 기준으로 삼아 4가지 문제점으로 분류한다. 문제점 분류를 위해 관람객이 콘텐츠 체험과정에서 나타낼 수 있는 얼굴 표정 3가지 종류와 제스처 5가지를 구분하였다. 실험에서는 입력된 비디오로부터 얼굴 및 손을 검출하기 위해 Adaboost algorithm을 사용하였고, mobilenet v1을 retraining하여 탐지모델을 생성 후 얼굴 표정 및 제스처를 검출했다. 이 연구를 통해 인터랙티브 디지털콘텐츠가 지니고 있는 문제점을 추론하여 향후 콘텐츠 개선과 제작에 사용자 중심 설계가 가능하도록 하고 양질의 콘텐츠 생산을 촉진 시킬 수 있을 것이다.

키워드: 감정인식(Emotion Recognition), 제스처 인식(Gesture Recognition), 인터랙션(Interaction), 디지털콘텐츠(Digital Contents)

I. Introduction

최근 가상증강현실 등의 4차 산업혁명 기술의 발전과 사람들의 문화 예술에 대한 수요가 높아짐에 따라 박물관이나 미술관 등의 전시공간에 인터랙티브 디지털콘텐츠의 보급이 활성화되고 있다. 그러나 현재는 초기 콘텐츠 제작과 전시 환경을 구축하는 방법에만 집중되어 있고, 구축된 콘텐츠의 개선 및 교체에 필요한 연구는 미흡한 상황이다. 따라서 본 논문에서는 관람객의 얼굴 표정 및 제스처 인식을 기반으로 인터랙티브 디지털콘텐츠의 문제점을 추론하는 시스템을 제안한다. 관람객이 콘텐츠와 상호작용을 하는 과정에서 나타나는 얼굴 표정과 특정한 행동 패턴을 시스템화 하여 구축된 콘텐츠가 지니고 있는 문제점을 추론한다. 이를 통해 추후 콘텐츠 유지보수와 개발과정에서 사용자 중심의 설계가 가능하도록 하여, 양질의 콘텐츠 생산을 촉진하고 사용자에게 만족도 높은 콘텐츠를 제공하는 것을 목표로 한다.

II. Preliminaries

1. Haar Feature-based Cascade Classifier

Haar Cascade Classifier는 영상에서 특정한 형태의 객체를 찾는 데 사용하는 대표적인 분류기로 Viola & Jones[1]에 의해 처음 제안되었으며 Rainer Lienhart에 의해 발전하였다. Haar Feature기반 방법은 Fig. 1과 같이 2개 이상의 명암대비가 있는 사각형의 집합으로 구성되어 있다. 어두운색과 밝은색 부분의 각 픽셀 값을 합한 다음 합쳐진 두 영역 픽셀 값의 차를 이용하여 표현한다.

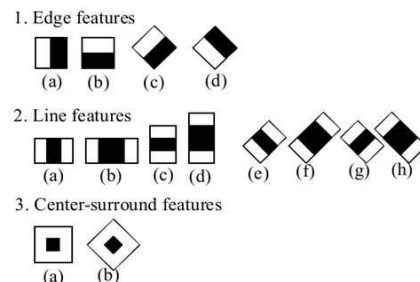


Fig. 1. Extended type of Haar-like features

또한 객체분류기는 AdaBoost 알고리즘을 이용하여 선형적인 다단계 분류기를 구성한다. AdaBoost 알고리즘은 약한 성능의 분류기 (weak classifier)를 순차적으로 학습시키고 연결하여 높은 검출 성능을 보이는 강한 성능의 분류기(strong classifier)를 생성하는 것이다. 약한 성능의 분류기를 하나씩 학습시키며, 이전의 학습된 분류기가 잘못 분류한 결과를 토대로 다음 학습에 사용하여 단점을 보완하도록 한다. 이로 통해 오분류한 데이터에 더 가중치를 두어 학습하여 강분류기를 생성해낼 수 있다.

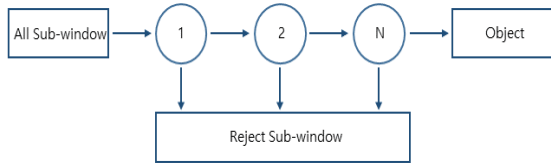


Fig. 2. Concept of AdaBoost Algorithm

III. The Proposed system

관람객의 얼굴 표정 및 제스처 인식을 기반으로 한 인터랙티브 디지털콘텐츠의 문제점 추론 시스템은 Fig. 3과 같다. 먼저 카메라가 관람객을 인식하면 콘텐츠 재생과 함께 사용자 제스처 인식을 시작한다. 시스템은 관람객이 콘텐츠를 체험하는 과정 중 그만두고 떠나는 시점의 행동과 표정을 기준 하여 총 4가지 문제점으로 분류된다.

1. Case #1 콘텐츠 시작단계에서 이동하는 경우

Case 1은 콘텐츠 시작단계에서 관람객이 아무런 동작을 하지 않거나 콘텐츠 체험과정에 필요한 인터랙션이 아닌 전혀 관계없는 제스처를 시행한 후, 부정적인 얼굴 표정을 지닌 채로 자리를 이동하는 경우이다. 이 경우 콘텐츠에서 제공하는 UI가 관람객의 체험 목표를 달성하기에 적절하지 않거나 콘텐츠 체험방법에 관한 충분한 설명이 부족하여 목표 사용자집단을 이해시키는 과정에서 문제가 있다고 판단한다.

2. Case #2, 3 인터랙션 중간에 이동하는 경우

인터랙션 중간에 이동하는 Case 2와 Case 3은 크게 두 가지 행동 패턴으로 구분할 수 있다. 첫 번째는 관람객이 콘텐츠 시작부터 자리를 이동하기 전까지 같은 인터랙션 동작을 계속 반복할 때이다. 두 번째는 관람객과 콘텐츠 사이의 인터랙션이 원활히 되지 않아 이동 직전 똑같은 제스처를 수 회 반복한 후 부정적인 얼굴 표정을 지닌 채 이동하는 경우이다. 관람객이 첫 번째 행동 패턴을 보일 때는 콘텐츠의 단순한 인터랙션 구성이 관람객의 흥미를 떨어뜨려 사용자 만족감을 저하시키는 원인으로 판단한다. 관람객이 두 번째 행동 패턴을 보인 경우는 인터랙션의 반응성과 사용성이 미흡하다고 판단한다.

3. Case #4 콘텐츠가 종료된 후 이동하는 경우

마지막으로 Case 4와 같이 관람객이 콘텐츠가 종료된 후 이동하면 콘텐츠 체험과정에서 인터랙션에는 문제점이 없다고 판단한다. 이 경우 제스처 인식을 제외한 얼굴 표정 인식만을 통해 콘텐츠 체험 후 사용자 감정을 파악하여 콘텐츠의 흥미도 및 문제점을 추론할 수 있다.

Table 1. Problem Inference Result of Contents

Case	Inference Result
Case #1	Lack of UI and usage Explanation
Case #2, #3	Problem with Interaction
Case #4	Less interesting of Contents

IV. Experiment

본 논문에서 제안하는 시스템은 Window 환경에서 Python과 OpenCV 2.0을 이용하여 구현되었다. 얼굴 및 제스처 검출과 분류를 위해 앞서 설명한 Haar Feature-based Cascade Classifier와 Mobilenet v1을 retraining 하여 모델을 구성하였다.

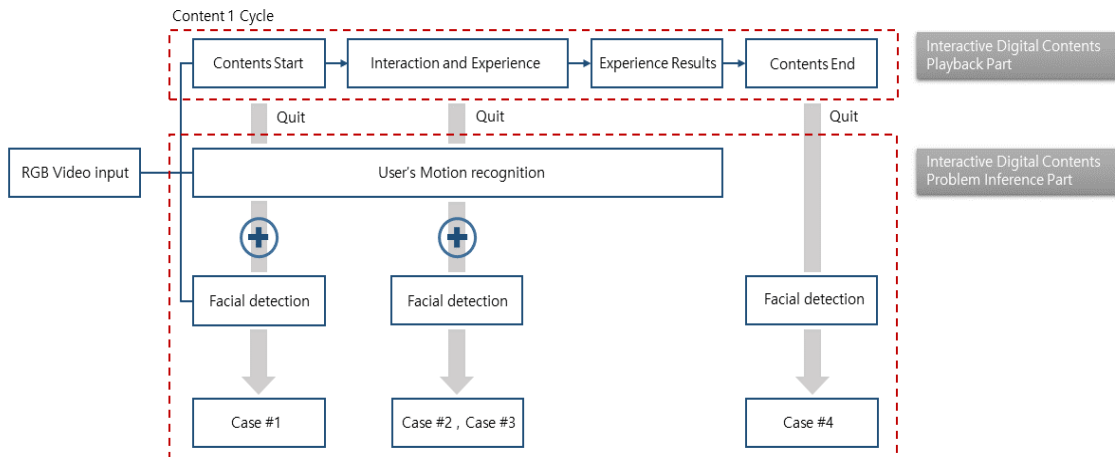


Fig. 3. System Architecture

1. 얼굴 표정 종류 정의 및 Dataset

콘텐츠를 체험할 때 나타날 수 있는 관람객의 표정은 러셀의 2차원 감정 공간[2]을 바탕으로 세 가지로 구분한다. retraining에 사용된 dataset은 The MPLab GENKI Database[3]와 웹스크레이핑을 통해 구성했다. 분류한 얼굴 표정의 종류와 data의 수는 Table 2와 같다. Fig 4는 훈련된 모델을 바탕으로 웹캠을 사용하여 얼굴 표정을 검출한 결과이다.

Table 2. Facial expression types and Dataset

Type of Expression	Number of Dataset
Positive	1187
Neutral	992
Negative	1120

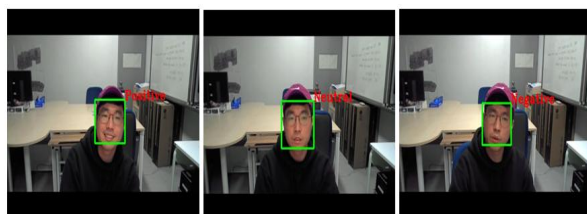


Fig. 4. Results of facial expression recognition

2. 제스처 종류 정의 및 Dataset

Table 3은 관람객이 콘텐츠와 인터렉션하는 과정에서 취할 수 있는 제스처의 종류와 data의 수를 나타낸다. 데이터는 직접 촬영한 이미지를 Image augmentation 하여 생성했다. Fig 5는 제스처를 인식한 결과 사진이다.

Table 3. Gesture types and Dataset

Type of Gesture	Number of Dataset
Spread	301
Point	256
OK-Shape	178
Fist	222
Grab	204

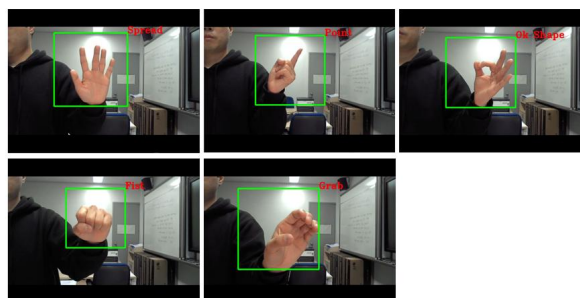


Fig. 5. Results of Gesture recognition

V. Conclusions

본 논문에서는 관람객 얼굴 표정 및 제스처 인식을 통한 인터랙티브 디지털콘텐츠의 문제점 추론 시스템을 제안했다. 관람객이 콘텐츠 체험을 그만두고 이동하는 시점을 기준으로 4가지 문제점으로 분류하고, 관람객이 체험과정에서 나타날 수 있는 얼굴 표정과 제스처를 정의하여 웹캠을 통해 탐지해보았다. 이 연구를 통해 향후 인터랙티브 디지털콘텐츠의 개선과 교체과정에서 사용자 중심 설계가 이루어지는 기초연구가 될 것이다.

ACKNOWLEDGEMENT

이 논문은 “2019 한국전통문화대학교 대학원 연구개발지원 사업”의 지원을 받아서 수행되었다.

REFERENCES

- [1] P. Viola and M. Jones “Rapid object detection using a boosted cascade of simple feature”, Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 511-518, 2001.
- [2] J. A. Russell, “A circumplex model of affect,” Journal of Personality and Social Psychology, Vol. 39, No. 6, pp. 1161-1178, 1980.
- [3] The MPLab GENKI Database, <http://mplab.ucsd.edu>