

왜곡 영상을 위한 효과적인 최소-최대 유사도(Min-Max Similarity) 기반의 영상 정합 알고리즘

허영진[†], 정다미^{**}, 김병규^{***}

An Efficient Image Matching Scheme Based on Min-Max Similarity for Distorted Images

Young-Jin Heo[†], Da-Mi Jeong^{**}, Byung-Gyu Kim^{***}

ABSTRACT

Educational books commonly use some copyrighted images with various kinds of deformation for helping students understanding. When using several copyrighted images made by merging or editing distortion in legal, we need to pay a charge to original copyright holders for each image. In this paper, we propose an efficient matching algorithm by separating each copyrighted image with the merged and edited type including rotation, illumination change, and change of size. We use the Oriented FAST and Rotated BRIEF (ORB) method as a basic feature matching scheme. To improve the matching accuracy, we design a new MIN-MAX similarity in matching stage. With the distorted dataset, the proposed method shows up-to 97% of precision in experiments. Also, we demonstrate that the proposed similarity measure also outperforms compared to other measure which is commonly used.

Key words: Image Processing, Image Matching, Similarity Measure, MIN-MAX Similarity, Distorted Images

1. 서 론

저작물이란 인간의 사상 또는 감정을 표현한 창작물로 어문, 음악, 연극을 비롯해 미술, 건축, 사진과 영상 등이 포함된다. 교과서를 비롯한 교육용 전자 출판물에서는 효과적인 정보 전달을 위해 다양한 저작물이 사용되는데 저작권 관련 법률에 따르면 고등학교 등 교육목적 상 필요한 교과용 도서에는 저작권자의 허락 없이도 공표된 저작물을 게재할 수 있다

[1]. 그러나 공표된 저작물만 사용할 수 있다는 조건에 따라 허락 없이 공표된 저작물을 사용하더라도 보상금을 지급해야 하는 규정이 존재하므로 원저작자의 요구가 있는 경우 보상금 지급이 이루어져야 한다.

인터넷의 보편화로 인해 사전에 공표된 저작물을 쉽고 간편하게 얻을 수 있는 경로는 늘어났지만 원저작자에게 정당하게 보상금을 지불하는 제도적 장치는 부족한 실정이다[2]. 이에 따라 최근 무분별한 저

* Corresponding Author: Byung-Gyu Kim, Address: (04310) Sookmyung Women's University, 100, Cheongpa-ro 47-gil Yongsan-gu Seoul, Republic of Korea, TEL: +82-2-2077-7293, FAX: +82-2-710-9704, E-mail: bg.kim@sookmyung.ac.kr

Receipt date: Sep. 24, 2019, Revision date: Nov. 28, 2019
Approval date: Dec. 16, 2019

[†] Dept. of IT Engineering, Sookmyung Women's University (E-mail: yj.heo@ivpl.sookmyung.ac.kr)

^{**} Dept. of IT Engineering, Sookmyung Women's University (E-mail: dm.jeong@ivpl.sookmyung.ac.kr)

^{***} Dept. of IT Engineering, Sookmyung Women's University

* This research project was supported by Ministry of Culture, Sports and Tourism (MCST) and from Korea Copyright Commission in 2019.

작물 사용에 있어 원저작자의 저작권을 보호하고 적절한 보상금 분배를 위한 인공지능 기반 저작권 관리 기술의 필요성이 대두되고 있다. 특히 목적에 맞게 저작물을 변형해서 사용하거나 여러 개의 저작물을 병합, 편집하여 사용하는 경우, 사용된 저작물의 각각의 저작자를 찾는 일은 더욱 어려워진다[3].

본 논문에서는 교과서에서 적용되는 사례처럼 다양한 왜곡으로 인해 변형된 영상 저작물의 객체를 원본에 가깝게 분리하고, 분리된 각 객체로부터 저작권 정보를 검색할 수 있는 영상 매칭 기술과 매칭에 사용되는 새로운 유사도 측정 기법을 제안한다. 본 논문에서 제안하는 왜곡 영상에 대한 영상 매칭 기술을 활용하여 교육적인 목적으로 사용하는 영상 등 저작물의 이용 및 정산 과정 자동화 기술을 발전시키고, 궁극적으로 합법적인 저작권 이용 활성화 및 저작권 침해를 사전에 예방하는 데 기여할 수 있다.

본 논문의 구성은 다음과 같다. 제2장에서는 관련된 연구들을 통해 적용 가능한 기술과 문제점에 대하여 언급한다. 제3장에서는 제안하고자 하는 영상매칭 관련 기법에 대해서 자세히 설명하고, 제4장에서는 제안된 기법의 영상 매칭 실험을 통하여 성능을 평가하고, 본 논문에서 제안한 기법의 차별성과 성능 결과를 통해 제5장에서 본 연구의 결론을 맺고자 한다.

2. 관련 연구

교과서에서 저작권 영상이 사용될 때 교육적 목적에 맞게 그 특성이 변형, 병합, 편집되는 경우가 많다. 이러한 왜곡 영상들에 대해서도 각각의 원본 영상을 검색하여 해당 저작권에 대한 정보를 추출할 수 있어야 한다. 이러한 경우 영상 매칭을 정확하게 수행하기 위해서는 왜곡에 강인한 기법을 적용해야 한다. 많은 제안된 영상 매칭 방법은 기본적으로 주요 특징점을 추출하고 이를 기반으로 유사도를 측정하는 것이 일반적이다[4-10].

영상 특징 점 기반 매칭 시에는 주로 영상을 대표할 수 있는 특징 점을 추출하여 이를 특징 기술자(keypoint descriptor)를 생성하고 두 특징 기술자를 정의된 유사도에 의해 정합도를 측정하는 방법이 사용된다. 가능한 모든 특징 점의 후보를 고려함으로써 매칭의 정확도를 높인 Scale Invariant Feature Transform(SIFT) [4] 알고리즘은 크기 변형이나 회

전 왜곡 등에 안정적인 성능을 보인다. SIFT 알고리즘을 개선하여 특징점의 분별력과 정확도 성능을 높인 알고리즘으로는 RootSIFT [5] 알고리즘과 weighted SIFT [6] 알고리즘이 있다. 또한, 키 프레임 특징들에 대해 가중치를 부여하여 검색 성능을 높인 연구 결과가 있다[20]. 하드웨어 설계 측면에서 영상 정합 효율을 높이기 위한 관련연구도 진행이 되었다[21-22]. Lee는 컨벌루션을 제거한 적은 연산량으로 필터를 설계하여 정합에 적합하도록 제시하였으며, 영상 스티칭(stiching)을 위해서 Lakshmi는 기존의 방법을 간소화하여 하드웨어에 적합하도록 설계하였다[21,22].

그러나 SIFT 계열의 알고리즘들은 계산이 복잡하여 수행 시간이 오래 걸린다는 단점이 있는데 이를 계산속도 측면에서 개선한 Speed Up Robust Feature(SURF) [7], Scale Invariant Corner Keypoints (SICK) [8] 같은 알고리즘들이 제안되었다. 성능은 뛰어나지만 특허권 문제로 상용 서비스에 사용하기 어려운 SIFT 알고리즘과 SURF 알고리즘을 대신하기 위해 OpenCV Labs에서 Oriented FAST and Rotated BRIEF(ORB) 알고리즘을 제안하였다[11]. ORB 알고리즘은 특징점 검출 알고리즘인 Features from Accelerated Segment Test(FAST) [9]와 특징 기술자 추출 알고리즘 Binary Robust Independent Elementary Features(BREIF) [10]을 결합하여 각도 변화에 대응할 수 있도록 하였다. 본 논문에서는 빠른 처리 시간을 위해 ORB 알고리즘을 적용하여 특징점을 추출하고 이후 매칭 과정에 사용한다. 본 연구에서는 ORB 알고리즘을 기반으로 좀 더 정확한 매칭 성능을 도출하고자 하였으며 매칭 기법을 새로운 방법으로 설계하였다.

3. 제안하는 효과적인 영상매칭 방법

본 논문에서 제안하는 영상 매칭 기법은 병합이나 편집과 같은 왜곡이 가해진 영상으로부터 객체를 분리하여 각 객체의 원본 영상과의 유사도 검출 단계를 거친다. 따라서 제안하는 방법은 크게 객체 추출 모듈과 영상 매칭 모듈로 분리될 수 있다.

Fig. 1에 제안된 영상 매칭 알고리즘의 전체적인 흐름도를 나타내었다. 먼저 입력된 영상을 구성하고 있는 다수의 객체를 분리한다. 분리된 객체 각각에서 ORB 알고리즘 [11]을 통해 특징점이 추출된다. 기존

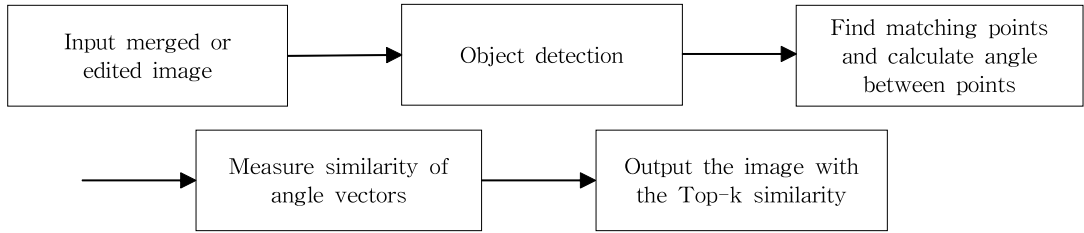


Fig. 1. Flowchart for overall image matching method.

의 BRIEF알고리즘은 벡터의 분산 값이 작기 때문에, 평균은 그대로 두고 높은 분산 값을 가지게 하는 알고리즘을 Rotated BRIEF라고 한다[17]. 입력 영상과 데이터베이스에 저장된 원본 영상 간의 특징 점 매칭을 통해 매칭된 30개의 특징점 쌍 사이의 제안하는 유사도 측정값의 비교를 통해 유사도 값이 큰 5개의 검색 결과를 출력한다.

3.1 회전 왜곡 복구 모듈

회전 왜곡은 픽셀의 기하학적 위치를 변화시키기 때문에 일반적인 특징점 매칭 기술을 적용할 시 오류율이 매우 높다. 따라서 이미지가 회전된 각도를 감지해 원본과 유사하도록 복구하면 정확도를 높일 수 있다. Fig. 2는 회전 왜곡에 대처하기 위한 복구 과정을 나타내고 있다. 먼저 팽창 연산을 통해 군집화된 이미지를 원래 이미지와의 차연산을 통해 윤곽을 드러낸다. 비트를 반전시켜 윤곽에 외접하면서 면적이 가장 작은 직사각형의 좌 상단 꼭지점 좌표를 계산한다. 이 좌표로부터 구해진 각도만큼 회전을 시키고 테두리 영역을 삭제하면 회전되기 이전보다 크기가 작은 이미지를 얻을 수 있다. 크기가 작은 이미지는 원본 이미지와 같은 지역적 특징을 가지고 있기 때문



Fig. 2. Recovery process for a rotated image.

에 이미지 매칭 및 검색에 용이한 형태가 된다.

3.2 객체 추출 모듈

회전 왜곡이 보상된 후 해당 입력된 영상을 구성하고 있는 객체를 분리하기 위해 객체 추출 모듈로 Mask R-CNN[12]을 사용한다. Mask R-CNN은 Faster RCNN [13]을 발전시킨 딥러닝 모델로써 각 픽셀이 객체에 해당하는 지 아닌지 판별하는 마스크를 추가하여 픽셀 단위의 조밀한 예측이 가능하며 Faster RCNN의 관심 영역 픽셀 조정 문제를 개선하여 객체 인식 성능 또한 향상시킨 모델이다. 본 논문에서는 약 5,000장의 데이터 셋으로 이루어져 Object detection, Segmentation, Keypoint detection과 같은 다양한 머신 러닝 분야에서 사용되는 MS-COCO [14] 데이터 셋으로 사전 훈련된 가중치를 이용하여 전이 학습을 통해 왜곡된 저작물에서 객체를 추출하는 데 적합한 모델로 Mask R-CNN을 학습시켰다.

Fig. 3과 Fig. 4에 실험에 사용한 왜곡된 데이터 셋을 나타내었다. Fig. 3과 같은 데이터 셋은 병합 왜곡 데이터 셋으로 분리할 수 있는 영상이 일부 면을 공유하며 단순히 붙어있는 형태를 보이는 데이터 셋을 의미한다. Fig 4의 (a), (b), (c)는 원본 영상으로부터 추출된 임의의 객체들이 모여 별도로 형성된 편집 왜곡 데이터 셋으로 특성에 따라 세부적으로 분류하였다. 편집 데이터 A는 원본 영상과 배경이 제거된 편집 영상의 합성으로 이루어진 데이터 셋이며, 편집 데이터 B는 배경이 제거된 편집 영상간의 합성으로 이루어진 데이터 셋이고, 마지막으로 편집 데이터 C는 배경이 제거된 단일 편집 영상이다.

객체 추출 모듈을 통해 왜곡된 데이터셋으로부터 객체를 추출한다. 객체 단위로 추출을 완료한 뒤, 추출된 각 객체를 입력으로 하여 영상 매칭 모듈을 통해 원본 영상과의 유사도를 비교하여 최종 검색 결과



Fig. 3. Example of the merged dataset.



Fig. 4. Examples of the edited dataset: (a) Edited Dataset A, (b) Edited Dataset B, and (c) Edited Dataset C.

를 도출한다.

3.3 영상 매칭 모듈

이전 절에서 객체 영상이 추출되면, 추출된 영상을 데이터베이스에 있는 원본 영상과 비교하는 단계를 거친다. 영상 매칭 모듈로 사용되는 ORB 알고리즘은 특징 점을 추출하기 위해 후보 픽셀을 중심으로 특정 반지름 위치에 놓인 픽셀들과의 밝기 값 차이를 비교하여 특징 점을 선정하고, 각 특징 점의 방향성을 부여하여 밝기나 회전 변형 등 왜곡에 강인하도록 설계된 알고리즘이다.

BRIEF(Binary Robust Independent Elementary Features) 알고리즘은 패치 내에서 가우시안 분포를 따르는 쌍들의 픽셀 밝기값을 비교하여 특징 기술자를 추출하는데 rBRIEF 알고리즘은 여기에 앞서 추출된 특징점의 각도에 대해 회전 행렬을 적용한 알고리즘으로 왜곡에 더 강인하다[17]. 수식 (1)은 이미지 하나에서 추출된 특징 기술자 벡터 결과이다.

$$D_i = \begin{pmatrix} v_{00} & \cdots & v_{031} \\ \vdots & \ddots & \vdots \\ v_{(k-1)0} & \cdots & v_{(k-1)31} \end{pmatrix} \quad (1)$$

여기서 i 는 이미지 번호를, k 는 특징점의 개수를 의미한다. 최종적으로 특징점 하나 당 32차원의 특징 기술자가 추출된다.

본 논문에서 제안하는 유사도 측정 방법은 입력된 영상으로부터 찾을 수 있는 모든 특징 점을 검색한 뒤 비교할 영상에서 추출된 특징 점들 간의 사이각을 계산한다. 수식 (2)은 입력 영상과 비교할 원본 영상의 특징 점집합을 나타낸다. 이때, X^q 는 특징 점의 x 좌표 집합이고 Y^q 는 특징 점의 y 좌표 집합이다.

$$I_{query} = [X^q, Y^q], X^q = \{x_1^q, x_2^q, \dots, x_n^q\}, Y^q = \{y_1^q, y_2^q, \dots, y_n^q\},$$

$$I_{original} = [X^o, Y^o], X^o = \{x_1^o, x_2^o, \dots, x_m^o\}, Y^o = \{y_1^o, y_2^o, \dots, y_m^o\}. \quad (2)$$

여기서 I_{query} 를 입력 영상의 특징점들의 집합이며 $I_{original}$ 를 원본 영상 특징 점들의 집합이라 할 때, 벡터 A와 B에 대한 정의는 수식 (3)와 같다.

$$A_i = \left| \operatorname{atan}\left(\frac{y_{i+1}^q - y_i^q}{x_{i+1}^q - x_i^q}\right) - \operatorname{atan}\left(\frac{y_{i+2}^q - y_{i+1}^q}{x_{i+2}^q - x_{i+1}^q}\right) \right| \quad (3)$$

$$B_i = \left| \operatorname{atan}\left(\frac{y_{i+1}^o - y_i^o}{x_{i+1}^o - x_i^o}\right) - \operatorname{atan}\left(\frac{y_{i+2}^o - y_{i+1}^o}{x_{i+2}^o - x_{i+1}^o}\right) \right|.$$

위 식에서 A_i 는 입력 영상의 매칭 점들의 사이 각 벡터, B_i 는 원본 영상의 매칭 점들의 사이 각 벡터를 구하는 식이며 이 벡터를 기준으로 다음 단계인 유사도를 측정한다.

다음과 같이 ORB 알고리즘을 통해 검출한 입력 객체 영상과 데이터베이스 상의 원본 영상 각각의 특징 점에서 특징 점들의 사이각을 벡터 A, B를 정의하고 이 벡터들의 유사도를 측정한다. 매칭된 점들이 실제로 의미 있게 연결되어 있는 지 유사도를 평가하기 위한 기준으로 본 논문에서 제안하는 Min-Max 유사도 매칭 기법은 수식 (4)과 같다.

$$Similarity = \frac{\sum_{i=0}^{n-1} MIN(A_i, B_i)}{\sum_{i=0}^{n-1} MAX(A_i, B_i)} \quad (4)$$

여기에서 n 은 벡터 A_i, B_i 의 크기를 의미한다.

위와 같은 식을 따라 입력 객체 영상 A와 데이터베이스 상의 영상 B의 상위 특징 점이 완전히 일치한다면 유사도 값은 1을 기대할 수 있다. 제안하는 Min-Max 유사도 매칭 기법은 Cosine Angle Distance(CAD) [15], [18], [19]와 같은 기존의 유사도 매칭 기법과 비교했을 때 더욱 간단한 계산으로 변별력 있는 결과를 얻을 수 있다는 장점이 있다. Min-Max 유사도 매칭 기법을 통해 측정된 유사도를 기준으로 데이터베이스 상의 영상 중에서 입력 객체 영상에 대해 가장 높은 유사도를 보이는 것들을 최종 매

칭 결과로 출력한다.

3.4 매칭 결과 출력

회전 및 밝기 변화, 크기 변화 등 왜곡이 편집 왜곡 등과 함께 존재하더라도 앞에서 언급한 제안된 기법으로 보정 후 Mask-RCNN을 적용하여 제안된 Min-Max 유사도 매칭 기법으로 최종 검색을 통해 결과를 도출하도록 하였다. 편집, 병합된 영상에 Mask-RCNN을 적용하여 객체 단위로 분리 한 다음, ORB 알고리즘으로 원본 영상과 특징 점들의 사이각을 비교하는 제안된 유사도 측정값을 활용한 결과는 Fig. 5와 같다. Fig. 5는 유사도 값이 가장 높은 5개의 영상을 출력한 결과로써 객체와 배경이 합성된 영상 (a)에서 같은 영상의 유사도 값은 약 0.964810이며 다른 영상과는 모두 유사도 값이 0.42 이하이다. 온전한 영상 두 개가 합성된 (b)의 경우에도 같은 또는 유사 영상의 경우 유사도 값이 0.992885이며 다른 영상과의 유사도 값은 0.45를 넘지 않는다. 즉 제안된 Min-Max 유사도 기반의 매칭 방법은 매우 변별력이 있음을 알 수 있다.

4. 실험 결과

4.1 실험 데이터 셋

본 논문에서 제안한 특징점 추출 및 매칭 기법이 교육 저작물에서 흔히 볼 수 있는 밝기, 회전, 크기

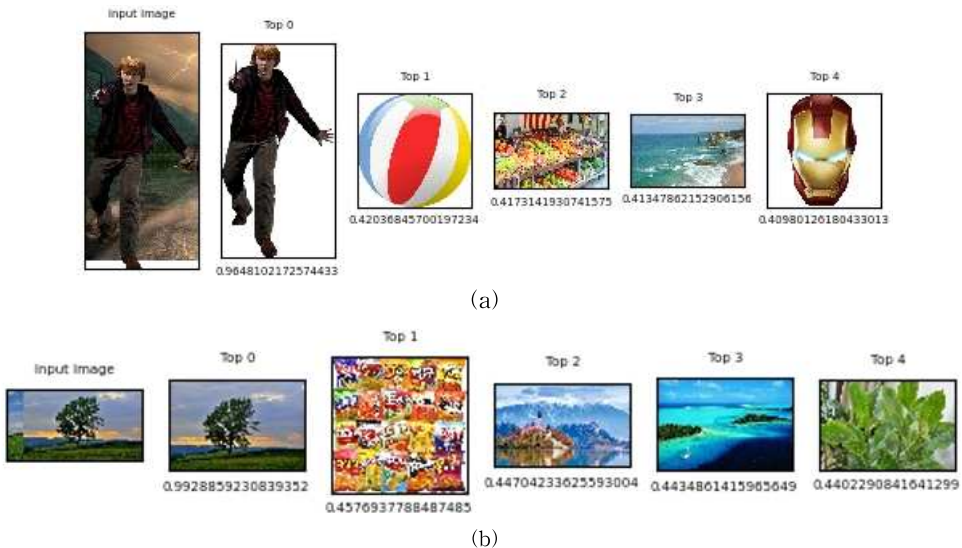


Fig. 5. Results of Top-5 ranked output: (a) Edited Image and (b) Merged Image.

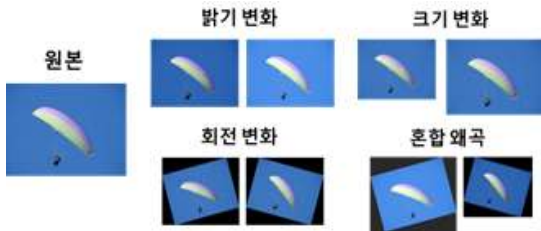


Fig. 6. Image examples with distortion of rotation, illumination change, and size change.

및 이를 혼합한 왜곡에 강인하다는 것 증명하기 위해 6가지 왜곡(Brighter, Darker, Bigger, Smaller, Bigger-brighter, Smaller-darker)을 적용한 변형 데이터 셋을 사용하여 실험하였다. Brighter 데이터 셋은 HSV 컬러공간으로 변환 뒤 V채널의 값을 40 증가시킨 데이터 셋이고, Darker 데이터 셋은 감마값을 0.85로 한 감마보정을 적용한 데이터 셋이다. Bigger, Smaller 데이터 셋 각각은 이미지 사이즈를 10% 증가, 감소시킨 데이터 셋이며 마지막으로 밝기와 크기 왜곡을 혼합한 데이터 셋에 시계방향, 반시계방향으로 회전 왜곡을 더한 데이터 셋이 Bigger-brighter, Smaller-darker 데이터 셋이다. Fig. 6은 이러한 왜곡 데이터 셋에 대한 예시이다. 특히 회전 영상은 보정 과정을 거친 후 다음 단계인 객체 추출 모듈로 입력이 된다.

4.2 Mask R-CNN 객체 추출 결과

객체 추출 모듈은 입력된 영상을 구성하고 있는 객체를 분리하는 것을 목적으로 한다. 병합 왜곡 데이터 셋, 편집 왜곡 데이터 셋에 맞게 전이 학습된 Mask R-CNN 객체 추출 모듈은 Fig. 7과 같이 입력 영상을 구성하고 있는 객체를 분리한다. 실험은 모두 269개의 병합 영상, 108개의 편집 영상으로 실행하였고 Mask-RCNN 수행을 통해 분할한 객체 수를 기반으로 병합영상은 563개, 편집영상은 172개의 영상으로 분할되어 졌다.

그러나 Fig. 8과 같이 실제 영상이 병합, 편집과 같은 왜곡이 없는 순수한 단일 영상이라도 객체 추출 모듈을 통해 영상 내에서 분리 가능한 객체가 있다면 다수 개의 객체가 추출될 수 있다. 이렇게 추출된 각 객체는 동일하게 데이터베이스 상의 영상과 매칭 과정을 수행한다. Mask R-CNN 객체 추출 모듈을 통해 객체가 지나치게 세분화되어 추출되는 경우 참-부정(True Negative: TN) 응답이 많아져 상대적으로 Precision을 낮추는 요인이 될 수 있다.

4.3 유사도 측정 방법 비교

최근 관련 연구 [18], [19]에서 Cosine Similarity를 이용한 매칭 방법이 연구되었으나, 본 논문에서는 원본 이미지와의 유사도 기반 향상된 변별력을 얻기



Fig. 7. Results of the object detection using Mask R-CNN.



Fig. 8. Examples of some over-segmentation results.

Table 1. Matching Performance Comparison of Cosine Similarity with Min-Max Similarity

Data type	Methods			
	Cosine Similarity[18]		Min-Max Similarity(Our)	
	TP-data Mean	FN-Top 1 Mean	TP-data Mean	FN-Top 1 Mean
Edited Data	0.957891	0.760928	0.930478	0.470727
Merged Data	0.974706	0.791545	0.954810	0.499616

Table 2. The accuracy performance of the proposed method

Data type	Proposed Method
	Accuracy [%]
Edited Data	87.20
Merged Data	96.98

위하여 Cosine Similarity가 아닌 Min-Max Similarity를 제안하였다. 특징 점 사이의 사이각을 의미하는 벡터 A와 B의 유사도를 측정하는 방법으로 기존 Cosine 유사도와 본 논문에서 제안하는 Min-Max 유사도를 비교하였다. 두 방법 모두 영상 매칭의 정확도 측면에서는 우수한 유사도 측정 기준이 되었으나, Cosine 유사도는 참-긍정(True Positive: TP)인 경우와 거짓-부정(False Negative: FN)인 경우의 유사도 차이가 크지 않아 임계치를 설정하기 어렵고 변별력이 높지 않음을 확인할 수 있었다.

Table 1에서 기존 Cosine 유사도와 Min-Max 유사도의 비교를 통한 측정 결과를 나타내었다. Table 1에서 TP-data Mean은 참-긍정(True Positive)인 경우에 나타나는 유사도의 평균값을, FN-top1 Mean은 거짓-부정(False Negative)인 경우에서도 가장 유사도가 높을 때의 값들을 평균한 값, 즉 정답이 아닌 것들 중 유사도가 가장 높은 것들의 평균값을 의미한다. Min-Max 유사도가 Cosine 유사도보다 참-

긍정(True Positive)값이 약 0.02정도 낮게 측정되는 데, 이는 특징 점들의 사이각이 모두 양수여서 유사한 방향 벡터 값을 갖기 때문이다. 따라서 Cosine 유사도는 참-긍정(True Positive) 데이터뿐만 아니라 거짓-부정(False Negative)인 경우에도 높은 유사도 값을 갖는다. 이 실험을 통해 정답과 정답이 아닌 것들의 유사도 차이를 크게 하는 Min-Max 유사도 측정 방법이 더욱 변별력이 있다는 것을 확인할 수 있다.

4.4 영상 매칭 결과

추출된 객체 영상을 입력으로 하여 영상 매칭 모듈을 수행하고 3.32절에서 제안한 유사도 매칭 기법으로 유사도를 측정하여 영상 매칭 정확도를 분석하였다. 영상 매칭 결과는 유사도가 가장 높은 5개의 영상을 출력해 주었으며, 임계치 0.6을 넘은 경우 원본 영상이 존재할 경우 맞았다고 측정하였다. 실험 결과는 Table 2와 같다. 편집 왜곡 영상 데이터 셋보다 병합 왜곡 영상 데이터 셋의 정확도가 더 높게 나타났다. 이는 편집 데이터 셋에서 객체 추출을 한 결과, 편집된 객체의 크기가 지나치게 작거나 편집된 영상 내에서 배경과의 구분이 잘 안가는 경우가 발생하여 매칭 정확도를 저하시키는 요인으로 작용하였다. 편집 왜곡 데이터 셋에서 이와 같은 현상이 더 자주 발생하여 매칭 결과의 정확도가 병합 왜곡 데이터 셋보다 낮은 것으로 나타났다.

Table 3. The performance comparison with an exiting method

Precision (%)		ORB [18]	Proposed
Simple distortion	Brighter	92.5	93.8
	Darker	95.0	98.5
	Bigger	93.6	99.8
	Smaller	96.0	99.2
Mixed distortion	Bigger-brighter	86.2	92.0
	Smaller-darker	88.2	97.7
Avg. Accuracy		91.9	96.8

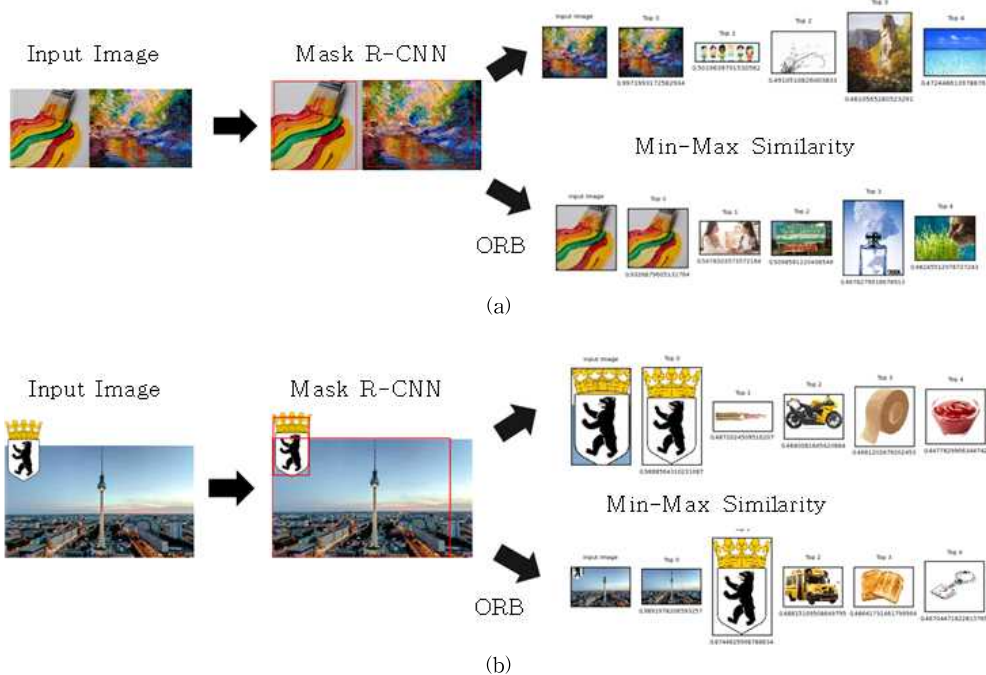


Fig. 9. Results of image matching: (a) Merged Image and (b) Edited Image.

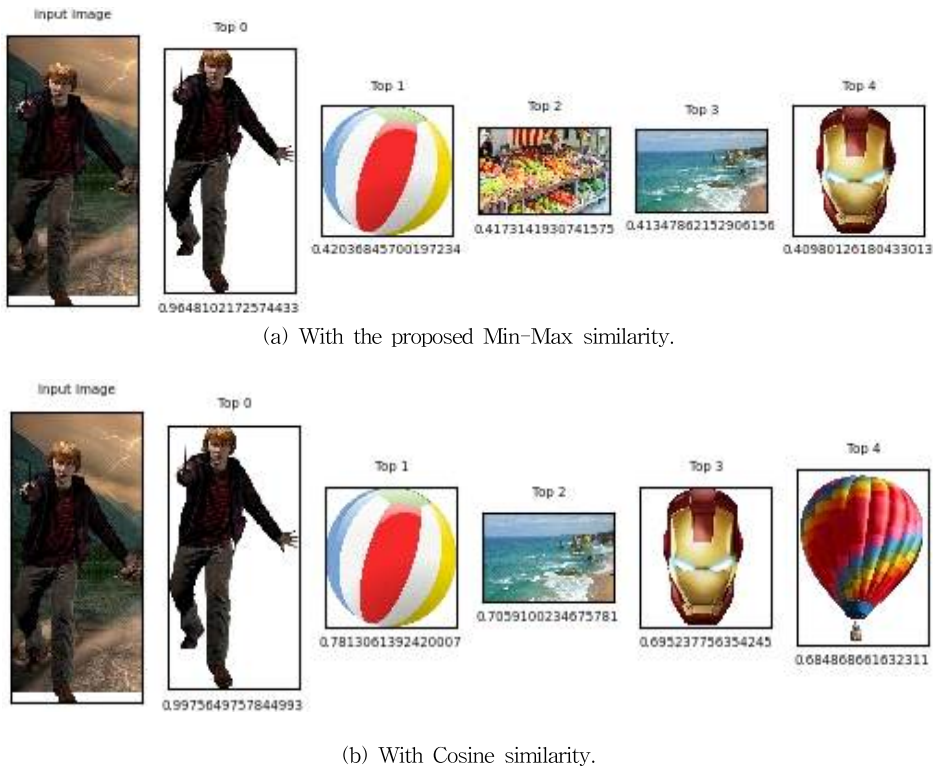


Fig. 10. Results of the proposed Min-Max Similarity and Cosine Similarity.



Fig. 11. Image retrieval results by the proposed matching algorithm.

Table 3에서 기존 ORB 알고리즘[18]과 제안한 방법의 성능을 비교하고 있다. 성능 비교를 위해 Precision을 정확도 지표로 사용하였다. 기존 ORB 알고리즘과 결과를 비교해 보면 제안된 방법이 회전 왜곡이 더해졌을 때 최대 10.7%까지 정확도가 증가한 것을 확인할 수 있다.

Fig. 9는 영상 매칭 결과를 전체적으로 보여준 그림이다. 첫 번째로 Mask R-CNN으로 객체를 분리시키고 각각 분리된 객체로부터 원본 영상과의 유사도를 비교하여 가장 높은 유사도를 가진 다섯 개의 영상을 출력하였다. 병합 데이터 (a)에서 같은 영상의 유사도 값은 각각 0.9971, 0.9326이고 편집 데이터 (b)에서 같은 영상의 유사도 값은 각각 0.9888, 0.9891이다.

Fig. 10에서는 Cosine 유사도와 제안된 Min-Max 유사도 기반의 영상 매칭 결과를 비교하였다. Fig. 10(a)와 같이 Min-Max 유사도로 측정된 결과 Top 1은 약 0.9648, Top 2는 약 0.4203이다. Fig. 10(b)에서는 Cosine 유사도로 측정된 결과를 보여준다. Top 1은 약 0.9975, Top 2는 약 0.7813로 나타났다. 해당 결과를 통해서 볼 수 있듯이 제안된 Min-Max 유사도는 Top 1과 Top 2의 차이가 0.5, Cosine 유사도를 적용 시는 0.2이므로 제안된 Min-Max 유사도가 더 변별력이 있음을 알 수 있다.

Fig. 11은 본 연구에서 제안한 기법에 의해 실제 입력에 대한 검색된 결과 영상을 보여 주고 있다. 결과에서 볼 수 있듯이 96% 이상의 정확도로 원본 영

상을 매칭하여 검출하였으며 회전, 크기 변화 등의 왜곡에 대해서도 매우 유사한 영상을 검색해서 찾아낼 수 있음을 관찰 할 수 있다. 이러한 성능 개선은 객체 기반의 정보가 추가됨으로써 단순히 특징 점에 의한 검색만으로는 어려운 부분을 어느 정도 보완할 수 있음을 보여준다고 할 수 있다.

5. 결 론

본 논문에서는 교육 저작물 등에서 사용되는 왜곡 영상에 대해 원본 영상과 매칭하기 위하여 Mask R-CNN을 이용한 객체 추출 모듈로 객체를 추출하고, ORB 알고리즘을 통해 특징 점을 추출했으며 특징 점 기반의 영상 매칭을 위해 Min-Max 유사도 측정 기법을 제안하였다. 특히 제안한 Min-Max 유사도가 기존 Cosine 유사도 측정 방법에 비해 안정적이고 특징에 대한 변별력 있음을 실험 결과를 통해 입증하였다. 영상 검색의 정확도 측면에서는 병합/편집의 경우 약 97%의 성능을 보였으며, 회전/크기 변화/밝기 변화 왜곡의 경우 약 96.8%의 정확도를 확인하였다.

향후에는 딥러닝을 이용하여 내용 기반 영상 검색 기술(Content-Based Image Retrieval)[16]을 추가하여 구조화된 객체 정보를 통하여 입력 영상과 일치하는 영상이 없더라도 구조화된 내용(객체) 기반으로 유사한 영상들을 검색 결과로 도출할 수 있는 시스템을 구축하는 것이 필요하다.

REFERENCES

[1] The Copyright Law, <http://www.law.go.kr/leg-islation/copyright-law> (accessed September 16, 2019).

[2] Y. Choi, J. Kim, J. Lee, Y. Lee, G. Hong, and B. Kim, "Efficient Object Classification Scheme for Scanned Educational Book Image," *Journal of Digital Contents Society*, Vol. 18, No. 7, pp. 1323-1331, 2017.

[3] D. Jeong, Y. Choi, and B. Kim, "A Study on Image Retrieval Algorithm for Scanned Education Books," *Proceeding of the Spring Conference of the Korea Multimedia Society*, pp. 158-161, 2019.

[4] D. Lowe, "Distinctive Image Features from Scale-invariant Keypoints," *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91-110, 2004.

[5] R. Arandjelovi and A. Isserman, "Three Things Everyone Should Know to Improve Object Retrieval," *Proceeding of Conference on Computer Vision and Pattern Recognition*, pp. 2911-2918, 2012.

[6] D. Jeong, J. Kim, Y. Lee, and B. Kim, "Robust Weighted Keypoint Matching Algorithm for Image Retrieval," *Proceeding of the 2018 the 2nd International Conference on Video and Image Processing*, pp. 145-149, 2018.

[7] H. Bay, A. Ess, T. Tuytelaars, and V. Gool, "Speeded-up Robust Features(SURF)," *Computer Vision and Image Understanding*, Vol. 110, No. 3, pp. 346-359, 2008.

[8] L. Bo, L. Haibo, and S. Ulrik, "Scale-invariant Corner Keypoints," *Proceeding of International Conference on Image Processing*, pp. 5741-5745, 2014.

[9] L. Yang, L. Lingshan, W. Lianghao, L. Dongxia, and Z. Ming, "Fast SIFT Algorithm Based on Sobel Edge Detector," *Proceeding of International Conference on Consumer Electronics, Communications and Networks*, pp. 1820-1823, 2012.

[10] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary Robust Independent Elementary Features," *Proceeding of European Conference on Computer Vision*, pp. 778-792, 2010.

[11] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An Efficient Alternative to SIFT or SURF," *Proceeding of IEEE International Conference on Computer Vision*, pp. 2564-2571, 2011.

[12] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," *arXiv Preprint arXiv:1703.06870*, 2017.

[13] S. Ren, K. He, R. Girshick, and J. Sun. "Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks," *arXiv Preprint arXiv:1506.01497*, 2015.

[14] T.Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, et al., "Microsoft COCO: Common Objects in Context," *Proceeding of European Conference on Computer Vision*, pp. 740-755, 2014.

[15] S. Amit, "Modern Information Retrieval: A Brief Overview," *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, Vol. 24, No. 4, pp. 35-43, 2001.

[16] J. Wan, D. Wang, S. Hoi, P. Wu, Ji. Zhu, Y. Zhang, et al., "Deep Learning for Content-Based Image Retrieval: A Comprehensive Study," *Proceeding of the 22nd ACM International Conference on Multimedia*, pp. 157-166, 2014.

[17] W. Huang, L. Wu, H. Song, and Y. Wei, "RBRIEF: A Robust Descriptor Based on Random Binary Comparisons," *Institution of Engineering and Technology Computer Vision*, pp. 29-35, 2013.

[18] L. Li, L. Wu, and Y. Gao, "Improved Image Matching Method Based on ORB," *Proceeding of the 17th IEEE/ACIS International Conference on Software Engineering*, pp. 465-468, 2016.

- [19] S. Eghbali and L. Tahvildari, "Fast Cosine Similarity Search in Binary Space with Angular Multi-index Hashing," *Proceeding of the IEEE Transactions on Knowledge and Data Engineering*, pp. 329-342, 2018.
- [20] K.W. Kim, "Improvement of Retrieval Performance Using Adaptive Weighting of Key Frame Features," *Journal of Korea Multimedia Society*, Vol. 17, No. 1, pp. 26-33, 2014.
- [21] S.J. Lee, K.S. Lee, and B.G. Kim, "Binary Image Based Fast DoG Filter Using Zero-dimensional Convolution and State Machine LUTs," *Journal of Multimedia Information System*, Vol. 5, No. 2, pp. 131-138, 2018.
- [22] L.K. Mudragada, K.S. Lee, and B.G. Kim, "Virtual Prototyping of Area-based Fast Image Stitching Algorithm," *Journal of Multimedia Information System*, Vol. 6, No. 1, pp. 7-14, 2019.



허 영 진

2019년 숙명여자대학교 IT공학과 학사
 2019년~현재 숙명여자대학교 IT공학과 석사과정
 관심분야 : 딥 러닝 기반 학습 알고리즘, GAN 구조, 지능형/인지기반 비디오 신호 처리 기술



정 다 미

2018년 숙명여자대학교 IT공학과 학사
 2018년~2019년 숙명여자대학교 IT공학과 석사과정
 관심분야 : 딥 러닝 기반 얼굴표정 인식기술, 비디오 신호처리, 지능형 비디오 서비스



김 병 규

1996년 부산대학교 전기공학과 학사
 1998년 한국과학기술원 전기 및 전자공학과 석사
 2004년 한국과학기술원 전기 및 전자공학과 박사
 2004년~2008년 한국전자통신연구원 선임연구원
 2009년~2015년 선문대학교 컴퓨터공학과 부교수
 2016년~2018년 숙명여자대학교 IT공학과 부교수
 2019년~현재 숙명여자대학교 IT공학과 교수
 2018년~현재 Journal of Multimedia Information System 편집위원장 및 한국멀티미디어학회 영문지 부회장
 관심분야 : 비디오 신호처리, 비디오 압축이론, 영상신호 처리 및 지능 시스템, 딥 러닝 기반 인식 알고리즘