

## 스케일에 강인한 LLAH 기반 문서 인식 알고리즘

이재하 박정주 박종일<sup>†</sup>  
 한양대학교 컴퓨터 · 소프트웨어학과  
 jhlee@mr.hanyang.ac.kr, jjpark@mr.hanyang.ac.kr, jipark@hanyang.ac.kr

## Scale-Invariant Document Detection Algorithm Based on LLAH

Jaeha Lee Jungjoo Park Jong-Il Park  
 Department of Computer and Software, Hanyang University

## 요 약

비슷한 코너의 모양을 가지는 다수의 글자가 포함된 문서 영상을 인식하는 일은 쉽지 않다. 일반적으로 성능이 우수하다고 알려진 SIFT 알고리즘은 코너를 기반으로 특징을 기술하는 알고리즘이기 때문에 각 글자가 비슷한 코너의 모양을 가지는 문서 영상 인식에서는 좋은 성능을 발휘하지 못한다. 반면, LLAH 는 각 단어의 크기를 알아내어 가우시안 필터와 이진화를 통해 단어를 하나의 점으로 나타내고 각 점과 점 사이의 기하 관계를 기술자로 표현하기 때문에 문서의 단어에서 점이 일관되게 추출된다면 좋은 인식 성능을 발휘한다. 그러나, 영상에서 단어의 크기를 알아내는 작업은 계산 측면에서 많은 비용을 필요로 한다. 이에 본 논문에서는 LLAH 를 사용하기 전에 반복적인 가우시안 필터와 이진화를 적용하여 단어의 크기를 알지 못하는 상황에서도 스케일에 강인하게 문서 영상을 인식할 수 있는 알고리즘을 제안한다.

## 1. 서론

영상의 국소 특징점 정보를 기술하고 기술된 정보를 바탕으로 영상을 구분할 수 있는 대표적인 알고리즘으로 SIFT 가 있다[1]. 일반적으로 SIFT 는 영상의 다양한 스케일과 회전에도 강건하게 수행되므로 객체 인식과 추적 등 다양한 분야에 활용되고 있다. 그러나, 코너를 기준으로 특징점 정보를 추출하는 SIFT 의 특성상 각 단어마다 유사한 코너 모양을 갖는 문서 영상의 경우 낮은 인식 성능을 보인다는 한계가 있다.

이를 해결하기 위해, 점 기반의 특징 기술자인 LLAH 가 제안되었다[2]. LLAH 는 각 단어를 하나의 점으로 표현하여 각 점들 간의 기하 관계를 특징으로 표현하는 알고리즘으로써 문서 영상 인식에 대단히 효과적이다. 각 점들 간의 기하 관계는 비조화비를 사용하여 호모그래피 변환에 강인한 특징이 있으며, 해시 구조로 기술자를 저장하고 있어 공간만 충분하다면 다른 알고리즘에 비해 비교적 빠르다는 장점이 있다[3]. LLAH 에서 단어를 점으로 표현하는 방법은 각 단어의 크기에 맞는 가우시안 필터와 이진화 처리를 이용하여 표현한다[4]. 그러나, 각 단어의 크기를 알아내는 것은 많은 비용을 필요로 하며 이에 대한 자세한 언급은 논문에서 언급되지 않았다. 계산 비용을 줄이기 위해 각 단어의 크기를 알아내는 작업을 생략할 경우 입력 영상의 스케일이 달라질 때마다 추출되는 점의 개수가 서로 다르게 추출되어 낮은 인식 성능을 보이게 된다.

본 논문에서는 LLAH 기반 반복적인 가우시안 필터와 이진화 처리를 통해 단어의 크기를 모르는 상황에서도

스케일에 강건하게 문서 영상을 인식하는 알고리즘을 제안한다.

## 2. 방법

그림 1 은 입력된 영상에서 단어의 크기를 알지 못하여 같은 단어의 형태를 가지는 서로 다른 스케일의 입력 영상에 고정된 가우시안 필터와 이진화 파라미터 값을 사용하였을 경우의 모습을 나타내고 있다.



그림 1. 입력 영상의 스케일에 따라 서로 다르게 추출되는 특징점의 모습

그림 1 에서 상단의 영상은 원본 문자 영상을 나타내며 하단의 영상은 원본 문자 영상에 가우시안 필터와 이진화를 차례대로 적용시킨 후 레이블링한 결과이다. 붉은색 점이 바로 레이블링된 각 요소의 중심 위치를 나타내는 특징점이다. 이 특징점들 간의 기하 관계가 LLAH 에 의해 계산된다. 그런데 그림 1 을 보면 입력 문자 영상의 스케일에 따라 서로 다른

<sup>†</sup> 교신저자

위치의 특징점들이 추출됨을 볼 수 있다. 이는 영상의 스케일에 따라 서로 다른 특징을 추출하여 계산하는 것이 되므로 추출된 각 특징점들 간의 기하 관계를 계산하는 LLAH의 특성상 낮은 인식 성능을 보이게 될 것을 의미한다. 그런데 다양한 파라미터를 가지는 가우시안 필터와 이진화를 쿼리 문서 영상에 반복적으로 적용한다면 레퍼런스 문서 영상에서 추출되었던 특징점들의 위치와 동일한 특징점들을 결국에는 추출해낼 수 있을 것이다. 이러한 아이디어를 바탕으로 본 논문에서는 다음과 같은 반복적인 알고리즘을 제안한다.

우선, 단어를 점 단위로 추출하기 위해 레퍼런스 영상에서 가우시안 필터와 이진화 작업을 통해 각 단어를 하나의 연결된 요소가 되도록 만든다. 이때, 가우시안 필터의 커널 사이즈를  $P_g$ 라 한다. 이진화는 적응적으로 수행되는데 평균 가중치를 사용하여 처리하며, 이진화에서 주변 픽셀을 사용하는 크기를 결정하는 파라미터를  $P_b$ 라 한다. 가우시안 필터와 이진화를 통해 얻어진 영상은 각 단어를 하나의 뭉쳐진 덩어리 형태로 나타내게 된다. 그 후, 각 덩어리의 중심을 계산하여 하나의 점으로 표현한다. 이렇게 획득한 각 점으로 LLAH 알고리즘을 수행하면 레퍼런스 영상의 특징 기술자가 해시에 저장된다.

쿼리 영상에서 점을 추출하기 위한 전처리 작업 역시 레퍼런스 영상을 처리하는 방식과 비슷하게 수행되나 가우시안 필터와 이진화 작업이 최대 4 번까지 반복된다는 차이가 있다. 쿼리 영상에서의 초기 가우시안 필터의 커널 사이즈를  $P_g/2$ 로 설정하고 초기 이진화 파라미터는  $P_b/2$ 로 설정한다. 그리고 LLAH 알고리즘을 통해 탐색을 수행한다. 만약, 일치하는 레퍼런스 영상을 찾지 못하면 가우시안 필터의 파라미터 값에  $P_g/3$ 을 더하고, 이진화 파라미터의 값에  $P_b/3$ 씩 더하여 다시 탐색을 수행한다. 이러한 반복은 최대 4 회까지 수행된다. 논문에서 제안하는 파라미터 값의 크기와 반복 횟수는 실험적으로 가장 나은 성능을 보이는 수치를 제안하는 것으로써 이보다 낮거나 더 클 경우는 성능 차이가 없었다.

### 3. 결과

각각 60 장의 레퍼런스 영상과 쿼리 영상에 대해 실험을 수행하였다. 단어의 크기를 알지 못하는 상황에서 입력 영상에 동일한 가우시안 필터와 이진화 파라미터를 이용하여 기존의 LLAH를 수행할 경우 80%의 인식률을 보였으며, 본 논문에서 제안하는 방법은 83.3%의 인식률을 보이며 3.3%가량 향상된 인식률을 기록하였다.

그림 2는 제안하는 방법의 처리 과정을 나타내고 있다. 그림 2의 좌측 상단의 영상은 레퍼런스 영상으로부터 추출된 점을 나타내며, 우측 상단의 영상은 쿼리 영상을 나타낸다. 2행과 3행의 영상들은 논문에서 제안하는 반복적인 방법에서 각 단계별로 추출되는 점의 모습을 나타내고 있다. 높은 파라미터 값을 가질수록 더 적은 양의 점이 추출되고 있음을 볼 수 있다. 우측 하단의 영상에서 쿼리 영상과 일치하는 레퍼런스 영상을 호모그래피 변환을 통해 증강시켜 표현하고 있다.

### 4. 결론

본 논문에서는 반복적인 가우시안 필터와 이진화를 통해



그림 2. 반복적인 알고리즘을 이용하여 문서 영상을 인식하는 과정

입력된 영상에서 단어의 크기를 알지 못하는 상황에서도 강인하게 문서 영상을 인식하는 알고리즘을 제안하였다. 또한, 실험을 통해 단어의 크기를 모르는 기존 상황 대비 3.3% 향상된 인식률을 획득하여 우수성을 증명하였다. 그러나, 입력되는 문서 영상에 글과 그림이 섞여 있는 경우에 글만 있는 경우보다 낮은 인식률을 보인다는 한계가 있었기 때문에 이를 해결하기 위한 추가 연구가 필요하다.

### 감사의 글

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2016년도 문화기술 연구개발 지원사업으로 수행되었음(R2015040004)

### 참고문헌

- [1] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, no. 2, pp. 91–110, 2004.
- [2] T. Nakai, K. Kise, and M. Iwamura, "Use of affine invariants in locally likely arrangement hashing for camera-based document image retrieval," *Proc. of DAS*, pp. 541–552, 2006.
- [3] H. Coxeter, *Projective geometry*, Springer, 2003.
- [4] D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*, Prentice Hall Professional Technical Reference, 2002.