MSER를 위한 획 너비 변환과 특징추출

[Stroke Width Transform and Feature Extraction for Maximally Stable External Regions(MSER)]

김민우[†] 오일석^{††} (Min-Woo Kim) (II-Seok Oh)

요 약 장면 텍스트 추출은 다양한 이미지 기반의 응용에 중요한 단서를 제공한다는 점에서 중요하다. MSER (maximally stable extremal regions)는 가장 우수한 영역검출 알고리금 중 하나로 전처리로서 텍스트 후보 영역을 찾는데 자주 사용된다. MSER 추출 후 각 영역이 텍스트인지판단하기 위해 획 정보를 추출해야 한다. 기존에 제안된 SWT(stroke width transform)는 에지를 기반으로 밝기 이미지를 획 너비 이미지로 변환하여 획 정보를 추출한다. 이방법은 획 정보 추출이 쉬우나 에지를 기반으로 동작하기 때문에 MSER에 바로 적용할 수 없다. 본 논문에서는 MSER의 획 너비 이미지를 생성하는 새로운 방법을 제안한다. 특히 제안하는 방법은 획의 너비 정보뿐만 아니라 영역이 평행한경계선을 갖는 정도를 나타내는 특징도 추출할 수 있으며, 이들 특징은 텍스트와 비텍스트를 구분하는 데 유용하다.

키워드: 획 너비 변환, MSER, 장면 텍스트 추출, 연결 요소, 특징 추출

Abstract Scene text extraction is crucial for diverse image-based applications due to the fact that the text provides contextual clues. Maximally stable extremal regions (MSER) are one of the most outstanding region detectors, and often used to detect text candidates in

the preprocessing step of scene text extraction. Once MSER is extracted, stroke information is needed to classify each candidate regions as text or background. The conventional stroke width transform (SWT) converts an intensity image into stroke width image based on edge points, and then extracts stroke information. This method is helpful to extract stroke information, but inappropriate for MSER to apply directly due to the fact that works based on edge points. In this paper, we propose a new method to make stroke width image for MSER. Especially, the proposed method can extract useful feature to classify a region into text or nontext about a degree that region is composed of parallel boundary besides stroke width information.

Keywords: stroke width transform(SWT), maximally stable extremal regions(MSER), scene text extraction, connected component, feature extraction

1. 서 론

디지털 이미지의 확산과 그에 따른 응용분야의 증가로 관련된 기술이 주목받고 있다. 특히 이미지에 포함된 텍 스트는 그러한 응용에 중요한 단서를 제공하기 때문에 중 요하다. 이러한 환경의 변화로 최근 10여 년간 장면 텍스 트 추출 문제는 많은 관심을 받으며 발전해왔다 [1-4].

전통적으로 장면 텍스트 추출 문제는 텍스처 기반 방법과 연결요소 기반 방법으로 분류할 수 있다[3]. 텍스처 기반 방법은 텍스트를 배경과 구분되는 특별한 텍스처로 보고 영상에서 그러한 텍스처를 찾는다. 일반적으로 슬라이딩 윈도 기법 등을 이용하여 검사 영역을 한정하고 그 영역에서 특징을 추출하여 미리 학습한 분류기로 텍스트의 존재 여부를 판단한다.

연결요소 기반 방법은 상향식 접근법으로 인접한 유사 연결요소를 그룹 짓고 텍스트가 가진 기하학적 특성을 이용하여 텍스트인 것을 찾는다. 일반적으로 영상에서 밝기와 색상 등이 유사한 인접 화소를 모아 연결요소를 만들고 글자 고유의 특성을 이용하여 글자가 아닐 가능성이 높은 것은 제거한다. 남겨진 연결요소는 유사한 인접 연결요소와 그룹 짓고 텍스트 라인의 기하학적특성을 만족하는지 평가한다.

장면 텍스트 추출 분야의 주목할만한 성과 중 하나로 Epshtein 등은 [5]에서 각 화소가 자신이 속한 획의 너비 값을 갖는 이미지 변환 SWT(stroke width transform)을 제안했다. 이 변환은 에지화소 p에서 그라디언트 방향으로 선을 그어 만나는 에지화소 q를 찾고, 두 에지화소가 서로 마주 보는 그라디언트를 갖는 경우 그 선 (p, q)을 획을 가로지르는 선으로 보고 수집한다. 수집한 선이지나간 자리에 선의 길이를 채워 넣음으로써 SW(stroke

bghunter@ibnu ac kr

** 종신회원 : 전북대학교 컴퓨터공학부/영상정보신기술연구소 교수

isoh@jbnu.ac.kr

(Corresponding author임)

논문접수 : 2013년 8월 20일 심사완료 : 2013년 10월 23일

Copyright©2014 한국정보과학회: 개인 목적이나 교육 목적인 경우, 이 저작 물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처 를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지: 컴퓨팅의 실제 및 레터 제20권 제1호(2014.1)

[·]이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임(2010-0010737)

[·]이 논문은 2013 한국컴퓨터종합학술대회에서 'MSER를 위한 획 너비 변환' 의 제목으로 발표된 논문을 확장한 것임

^{*} 학생회원 : 전북대학교 전자정보공학부

width) 이미지를 얻는다. 그 논문은 글자가 일정한 두께의 획으로 구성되었다는 사실에 기반하여 SW가 유사한화소를 모아 연결요소를 만들었다. 주목할 점은 SW 이미지에서 추출한 획(stroke) 정보를 텍스트를 배경과 구분하기 위한 중요한 특징으로 사용했다는 점이다. 이후다양한 논문이 획 정보를 글자를 찾고 텍스트 라인을형성하기 위한 기본 특징으로 사용하고 있다[5-11].

장면 텍스트 추출 분야의 또 다른 주목할만한 호름은 연결요소 추출에 MSER를 사용하는 것이다. MSER는 가장 우수한 영역 검출 알고리즘 중 하나로 주변에 비해 밝거나 어두운 영역 중 문턱치 변화에 안정적인 것들을 찾는다[12,13]. 텍스트는 일반적으로 밝기가 일정하고 배경과의 차이가 분명하다. 이러한 특성이 MSER와 잘 맞아떨어져 텍스트 추출 연구에서 연결요소를 찾기위해 MSER가 자주 사용되며 우수한 성능을 보이고 있다[8-11,14-16].

여기에서 발생하는 하나의 이슈는 MSER에서 휙 정보를 어떻게 추출하느냐이다. SWT는 에지 영상을 기반으로 동작하며 에지의 위치는 MSER의 경계와 일치하지 않는다. MSER의 경계에서 그라디언트 방향으로 선을 그으면 선의 방향이 부정확할뿐더러 일부는 획의 외부로 향한다. 따라서 MSER의 경계를 에지로 보고 SWT를 적용하는 데에는 무리가 있다. 다른 대안으로는 MSER와 별도로 입력 이미지에 SWT를 적용하여 SW이미지를 생성하고 MSER 상의 각 화소의 값으로 같은 위치의 SW 이미지의 화소 값을 취하는 것이다. 하지만 MSER의 경계에 해당하는 에지가 검출되지 않으면 SW이미지에 값이 존재하지 않게 된다. 결국 MSER를 기반으로 하는 알고리즘이 에지 이미지에 의존적이게 되는 문제가 있다.

2. 관련연구

Chen 등은 [9]에서 MSER를 기본 문자 후보로 사용하는 연결요소 기반 장면 텍스트 추출 방법을 제안했다. 추출된 MSER를 Canny 에지를 이용해 영역의 경계를 보강하고 규칙 기반의 필터링 및 그룹화 과정을 거쳐 텍스트 영역을 찾는다. 이 규칙들 중 획 정보가 중요한역할을 한다. 이 논문에서 MSER에서 SW 이미지를 구하기 위해 거리변환(distance transform)을 이용한다. MSER를 거리변환 하면 각 화소는 가장 가까운 배경화소와의 거리를 값으로 갖는다. 변환된 거리 이미지의 능선의 값이 획 너비의 절반에 해당하는 값이다. 이 논문에서는 능선의 값을 언덕 아래로 전파하여 SW 이미지를 만든다.

Li 등은 [11]에서 MSER를 기반으로 하는 연결요소 기반 장면 텍스트 추출 방법을 제안했다. Chen 등의 방 법 [9]과 차이점은 MSER를 보강하기 위해 대비를 강화한 이미지를 사용한 점과 획 정보를 구하는 방법 등이다. 이 논문에서는 MSER의 스켈레톤을 구하고, 스켈레톤 상의 각 화소에서 MSER의 배경까지의 거리를 구하여 스켈레톤-거리 이미지를 생성한다. 스켈레톤 상의화소들은 거리변환 이미지의 능선과 거의 일치한다. 따라서 스켈레톤-거리 이미지는 획 너비의 절반에 해당하는 값으로 채워진다.

Chowdhury 등은 [7]에서 Canny 에지를 연결요소 분석하여 후보 영역을 찾고 MLP(multi-layer perceptron)를 이용해 텍스트 영역을 구분하는 연결요소 기반 장면텍스트 추출 방법을 제안했다. MLP 학습을 위한 특징으로 획 정보가 사용된다. 이 논문에서는 후보 영역을이진화하여 연결요소를 얻고 연결요소의 거리변환 이미지를 구한다. 거리변환 이미지의 능선의 값을 얻기 위해지역-적응 이진화하고, 전경(foreground) 화소에 해당하는거리변환 이미지의 값만을 취하여 획 정보를 계산한다.

Yin 등은 [10]에서 MSER를 글자 후보로 취하고 규칙기반의 필터링과 그룹화를 거쳐 텍스트 후보 영역을 찾는다. 마지막으로 텍스트 후보 영역은 AdaBoost 분류기를통해 검증한다. 이 논문에서도 획 정보가 중요하게 사용된다. 이 논문에서는 MSER 상의 각 화소에서 방사형으로 4방향의 선을 긋고 이 중 배경까지의 거리가 가장 짧은 선의 길이를 획의 너비로 취하는 방법을 제안했다.

Yi등은 [6]에서 MSER 상의 각 화소에서 획의 방향을 Gabor 필터를 이용해 추정하고, 획의 방향과 수직으로 선을 그어 배경까지의 선의 길이를 획의 너비로 취한다.

획은 두께가 일정한 띠 모양의 연속된 영역으로 정의할 수 있다[5]. 즉 획의 중요한 특성은 두께가 일정하다는 것과 띠 모양이라는 것이다. 특히 후자의 띠는 마주보는 경계가 평행하다는 것으로 해석할 수 있다. 앞서언급한 방법 중 SWT를 제외한 나머지는 배경까지의거리, 즉 두께만을 염두에 둔다. 반면 SWT는 선의 끝에 존재하는 두 에기 화소가 마주 보는 그라디언트를가질 때만 획으로 인정한다. 이 연산에는 획이 평행한두 경계를 가진다는 가정이 포함되어 있다.

본 논문은 MSER의 경계에서 획을 가로지르는 선을 그어 SW 이미지를 구하는 SWT와 유사한 새로운 방법을 제안한다. 제안한 방법은 앞서 언급한 획의 두 가지특성을 모두 고려할 뿐만 아니라 기존 방법에서 구할 수 없었던 새로운 특징을 얻을 수 있다.

3. MSER에서 SW 이미지 생성 및 특징 추출

획의 두께를 측정하기 위해 획의 경계에서 획을 가로 지르는 선을 그어 선의 길이를 샌다. SWT에서는 이 선 의 방향을 그라디언트를 이용하여 결정했다. 하지만 MSER의 경계와 에지는 정확히 일치하지 않아 MSER 의 경계에서 그라디언트 방향이 부정확할 뿐만 아니라 획의 외부로 향할 수 있다.

우리는 이 문제를 해결하기 위해 그림 1과 같이 MSER의 경계에서 접선에 수직인 방향으로 선을 긋는 방법을 제안한다. MSER의 경계화소 p에서 접선에 수직인 방향으로 선을 그어 만나는 다른 경계화소 q를 찾고 q의 접선이 p의 접선과 평행이면 그 선 (p, q)를 수집한다. 이때 점 p에서 시작하는 선의 방향은 근사화하여 구한다. 점 $p(p_x, p_y)$ 를 중심으로 n*n 마스크 내부에 있는 연결요소에 포함된 화소들의 집합을 M이라 할 때 선의 방향 $d(d_x, d_y)$ 는 식 (1)과 같이 구한다. c와 p는 2차원 공간상의 좌표이다. (본 논문에서 n=5를 사용했다.)

$$d(d_x, d_y) = c - p,$$

$$c(c_x, c_y) = \sum_{m(x, y) \in M} {m \choose |M|}$$
(1)

수집한 모든 선에 대해 그것이 지나간 화소에 선의 길이를 채워넣는다. 이때 이미 값이 채워져 있는 화소는 더 작은 값을 취한다. SWT에서와 마찬가지로 획의 교 차점에서 값의 부정확함을 보완하기 위해 한 번의 처리 과정을 더 거친다. 수집한 각 선에 대해 선이 지나간 화소들의 중앙값 m을 찾고 이보다 큰 값을 m으로 대체한다. 전체 처리 과정은 Epshtein 등의 SWT와 같으므로보다 상세한 설명은 [5]에서 확인할 수 있다.

기존 방법들이 획에서 특징을 추출하기 위해 사용한 정보는 영역 내 획 너비의 평균, 중간값, 편차이다[5-11]. 일반적으로 글자는 일정한 두께의 획으로 구성되어 있고, 획은 평행한 경계를 갖는 구간이 많은 비중을 차지한다. 우리는 유용한 특징을 하나 더 소개한다. 그것은 전체 경계 화소에서 그은 선 중에서 쌍을 이루는 선의 비율로, 이 값이 클수록 영역은 평행한 경계에둘러싸여 있다는 의미이다. 즉 이 값이 클수록 텍스트연결요소일 가능성이 크다.



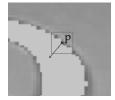


그림 1 연결요소에서 경계의 접선에 수직인 방향으로 선을 그어 획의 두께를 잰다.

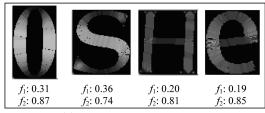
Fig. 1 Shooting a ray in perpendicular direction to the tangent to calculate stroke width of a connected component

4. 실험 및 결과

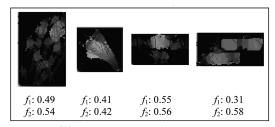
실험을 위해 텍스트와 비텍스트의 연결요소를 수집한다. 이미지는 ICDAR 2011[2] 표준 데이터 집합의 훈련 집합을 사용하고 연결요소는 MSER를 이용해 추출한다. 연결요소는 텍스트인 것과 텍스트가 아닌 것으로 나눈다. 이때 텍스트인 것 중에서 심하게 오염된 연결요소는 제외한다. 수집한 연결요소는 텍스트가 총 8,218개, 비텍스트가 총 17,129개이다. 수집한 모든 연결요소에 대하여 제안한 방법으로 SW 이미지를 생성하고 다음 두 가지 특징을 추출했다(그림 2).

특징 fi은 획 너비의 변화의 심한 정도를 표현한다. 그림 3은 텍스트와 비텍스트 부류 각각에 대한 특징 fi의 분포이다. 텍스트 연결요소가 비텍스트 연결요소보다일반적으로 값이 작은 것을 확인할 수 있다. 이 값이 문택치(보통 0.5를 사용)를 넘는 연결요소는 비텍스트로보고 제거하는 경우가 많다[5,9,11]. 특징 f₂는 연결요소가 평행한 경계를 가지는 정도를 표현한다. 그림 4는 두부류 각각에 대한 특징 f₂의 분포이다. 텍스트가 비텍스트보다 일반적으로 큰 값을 가지는 것을 확인할 수 있다. 비텍스트 중 값이 1 부근에 모여있는 것은 배경 중작은 사각형 모양의 연결요소들로서 이런 부류의 연결요소는 대부분의 영역에서 평행한 경계를 갖는다.

f₁: 획 너비의 표준편차 / 획 너비의 평균 f₂: 수집한 선의 수 / 경계 화소의 수



(a) Text connected components



(b) Nontext connected components

그림 2 텍스트 연결요소와 비텍스트 연결요소의 SW 이미 지와 특징이다. 획의 너비가 넓을수록 밝게 표시했다.

Fig. 2 Stroke width image and feature value of text and nontext connected components. Brighter pixels represent wider stroke width

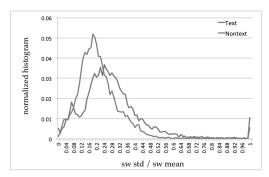


그림 3 특징 f_1 의 분포 Fig. 3 Distribution of f_1

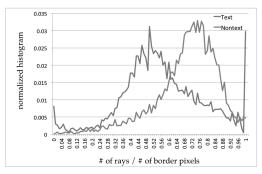


그림 4 특징 f_2 의 분포 Fig. 4 Distribution of f_2

그림 3과 4를 비교하면 특징 f_2 가 f_1 보다 두 부류에 대해 분포가 분리된 것을 확인할 수 있다. 이 사실을 바탕으로 텍스트/비텍스트 분류 문제에서 특징 f_2 가 더 큰 분별력을 가질 것으로 예상할 수 있다.

다음 실험은 앞서 수집한 텍스트 연결요소와 비텍스트 연결요소를 Random Forest[17,18] 분류기를 이용하여 학습시켰다. 특징은 앞에서 언급한 2가지 특징과 일반적으로 자주 사용되는 다음의 7가지 특징을 합하여총 9개의 특징을 사용했다.

 f_3 : 연결요소의 너비 / 높이

f4: 연결요소의 화소 수 / (너비*높이)

fs: 경계화소의 수 / ((너비+높이)*2)

f₆: 획 너비의 중앙값 / ((너비+높이)/2)

 f_7 : Border energy [11]

f8: 수직방향의 B/W transition의 분산

fg: 수평방향의 B/W transition의 분산

본 실험과 같은 2부류 분류 문제에서 Random Forest의 출력은 텍스트 부류에 대한 신뢰도 값으로 이용할 수 있다. 텍스트와 비텍스트를 분류하는 신뢰도 기준에 따라 분류기의 정확률(precision)과 재현율(recall) 이 달라진다. 그림 5는 학습 후 학습집합에 대해 신뢰도

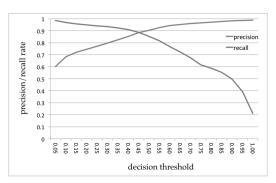


그림 5 신뢰도 문턱치에 따른 정확률과 재현율 그래프 Fig. 5 Precison and recall graph according to the confidence threshold

표 1 Random Forest가 계산한 특징 $f_1 \sim f_9$ 에 대한 중요도 Table 1 Importance of $f_1 \sim f_9$ features produced by Random Forest

Feature	Importance	Feature	Importance
f_1	0.09	f_6	0.11
f_2	0.14	f_7	0.10
f_3	0.07	f_8	0.08
f_4	0.08	f_9	0.14
f_5	0.19		

기준을 변화시켜가며 정확률과 재현율을 계산하여 표시한 것이다. 연결요소를 텍스트와 비텍스트로 구분하는 문제는 부류 간 유사한 샘플이 많아 분류가 어려운 문제에 속한다. 장면 텍스트 추출 문제에서 연결요소 분류는 전체 시스템의 초반부이며, 이 단계에서 비텍스트로 분류된 연결요소는 이후 그룹화 단계에서 배제된다. 따라서 연결요소 분류 시 정확률보다는 재현율을 우선하여 신뢰도 문턱치를 결정해야 한다. 따라서 문턱치는 0.1~0.2 정도의 낮은 값으로 설정하여 높은 재현율을 유지한다. 그렇더라도 전체 비텍스트 연결요소 중 70%가량을 걸러낼 수 있어 이후 처리에 도움이 된다.

Random Forest는 학습 과정에서 각 특징의 중요도를 계산한다. 이 값을 통해 각 특징의 기여도를 유추할수 있다. 표 1은 학습 결과 Random Forest가 계산한각 특징에 대한 중요도를 표로 나타낸 것이다. 휙 너비는 f_1 , f_2 , f_3 의 세 가지 특징과 관련되어 있다. 결과에서알수 있듯 제안하는 방법으로 구한 SW 정보가 연결요소를 텍스트와 비텍스트로 구분하는데 크게 기여함을확인할수 있다. 특히 새롭게 추가된 특징 f_2 가 f_1 보다중요도가 높은 것을 확인할수 있다.

5. 결 론

본 논문은 MSER에서 SW 이미지를 얻기 위한 새로

운 방법을 제안했다. SWT는 에지 이미지를 기반으로 동작하며 MSER의 경계와 에지 화소가 일치하지 않아바로 적용할 수 없다. 대신 우리는 MSER의 경계화소에서 접선의 수직 방향으로 선을 그어 획의 너비를 재는 새로운 방법 제안하였다. 그뿐만 아니라 이 과정에서 얻을 수 있는 마주 보는 경계화소의 비율은 텍스트와비텍스트를 구분하는 유용한 특징으로 사용될 수 있다. 실험을 통해 제안하는 SW 변환이 잘 동작하며 제안하는 특징이 텍스트와 비텍스트를 구분하는 분별력을 가짐을 확인할 수 있었다.

References

- S. M. Lucas, "ICDAR 2005 text locating competition results," Proc. of ICDAR, pp.80-84, Aug. 2005.
- [2] A. Shahab, F. Shafait, and A. Dengel, "ICDAR 2011 robust reading competition challenge 2: reading text in scene images," *Proc. of ICDAR*, pp.1491–1496, Sept. 2011.
- [3] K. Jung, K. Kim, A. Jain, "Text information extraction in images and video: a survey," *Pattern Recognition*, vol.37, no.5, pp.977-997, May 2004.
- [4] H. Zhang, K. Zhao, Y.-Z. Song, J. Guo, "Text extraction from natural scene image: a survey," *Neurocomputing*, vol.122, pp.310–323, Dec. 2013.
- [5] B. Epshtein, E. Ofek, Y. Wexler, "Detecting text in natural scenes with stroke width transform," *Proc. of CVPR*, pp.2963–2970, Jun. 2010.
- [6] C. Yi, Y. Tian, "Localizing text in scene images by boundary clustering, stroke segmentation, and string fragment classification," *IEEE Transactions* on *Image Processing*, vol.21, no.9, pp.4256–4268, Sept. 2012.
- [7] A. R. Chowdhury, U. Bhattacharya, S. K. Parui, "Scene text detection using sparse stroke information and MLP," Proc. of ICPR, pp.294–297, Nov. 2012.
- [8] C. Shi, C. Wang, B. Xiao, Y. Zhang, and S. Gao, "Scene text detection using graph model built upon maximally stable extremal regions," *Pattern Recognition Letters*, vol.34, no.2, pp.107-116, Jan. 2013.
- [9] H. Chen, S. S. Tsai, G. Schroth, D. M. Chen, R. Grzeszczuk, B. Girod, "Robust text detection in natural images with edge-enhanced maximally stable extremal regions," *Proc. of ICIP*, pp.2609–2612, Sept. 2011.
- [10] X. Yin, X.-C. Yin, H.-W. Hao, K. Iqbal, "Effective text localization in natural scene images with MSER, geometry-based grouping and AdaBoost," *Proc. of ICPR*, pp.725–728, Nov. 2012.
- [11] Y. Li, H. Lu, "Scene text detection via stroke width," Proc. of ICPR, pp.681–684, Nov. 2012.
- [12] J. Matas, O. Chum, M. Urban, T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal

- regions," *Image and Vision Computing*, vol.22, no.10, pp.761–767, Sept. 2004.
- [13] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A comparison of affine region detectors," *International Journal of Computer Vision*, vol.65, no.1–2, pp.43–72, Nov. 2005.
- [14] C. Merino-Gracia, K. Lenc, M. Mirmehdi, "A head-mounted device for recognizing text in natural scenes," *Proc. of CBDAR*, pp.29–41, 2011.
- [15] L. Neumann, J. Matas, "Text localization in realworld images using efficiently pruned exhaustive search," Proc. of ICDAR, pp.687–691, Sept. 2011.
- [16] H. Il Koo, D. H. Kim, "Scene text detection via connected component clustering and nontext filtering," *IEEE Transactions on Image Processing*, vol.22, no.6, pp.2296–2305, Jun. 2013.
- [17] L. Breiman, "Random forests," Machine Learning, vol.45, no.1, pp.5–32, Oct. 2001.
- [18] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, 2nd Ed., pp.587–604, Springer, New York, 2009.