

# 차량 전방 사진을 통한 혼잡도 측정

이웅규<sup>0</sup> 남종호

서강대학교 컴퓨터공학

taewoongkyu@naver.com, jhnang@sogang.ac.kr

## Measuring Traffic Congestion of Road through the Vehicle-front Image Analysis

Woongkyu Lee<sup>0</sup> Jongho Nang

The Sogang University

### 요 약

나날이 늘어가는 교통혼잡을 운전자가 미리 알 수 있도록 차량 안에서 찍은 전방사진을 분석하여 도로의 혼잡도를 분석할 수 있도록 하는 것이 이 논문의 목적이다. 대부분의 연구들이 전방의 사진에서 차량의 특징을 분리하여 혼잡도를 판단하도록 하는데 비하여 이 논문에서는 전체적인 도로상의 질감을 이용하여 혼잡도를 판단하도록 한다. 전방의 사진들은 최근 영상 분석에 많이 쓰이는 Bag-of-Words 방식을 이용하여 분석한다. Bag-of-Words의 방식은 이미지들의 특징을 추출하여 학습모델로 나타내고, 모델을 이용하여 테스트 데이터(이미지)들의 입력을 받아 결과를 총 3가지의 클래스(매우 혼잡, 혼잡, 원활)로 분류하여 사용자에게 간단한 도로 혼잡 정보를 전달할 수 있도록 해준다.

### 1. 서 론

오늘날의 교통상황은 소득의 증가와 여가에 대한 관심이 높아지면서 지속적으로 차량이 증가하여 매우 혼잡한 경향을 보이고 있다. 그래서 영상 처리 기법을 기반으로 하여 교통의 혼잡도를 측정하는 시스템의 개발이 활발하게 연구되고 있다. 만약, 혼잡도 검출에서 사람이 직접 혼잡도를 관측하고 분석을 한다면, 시간과 비용이 많이 들어가기 때문에 자동으로 혼잡을 판단해 줄 수 있는 시스템이 필요하다.

기존의 연구들을 살펴보면, 도로에서 일정 시간 동안 그 구간을 지나는 차량의 수를 계산하는 방법이나[1], 영상에서 그림자를 제거한 움직이는 차를 검출하여 혼잡도를 계산하는 방법[2] 등이 있었다. 하지만 이런 방법들은 실시간성이 떨어지고, 여러 영상들이 요구되며, 배경에 민감하게 반응한다는 단점이 있다.

본 논문에서는 위의 문제점을 해결하고자 차량의 전방 카메라를 이용하여 자동으로 사진을 찍어서 한 장의 사진을 분석하여 혼잡 정보를 다른 사용자에게 바로 전달하는 방법을 제시한다. 사진을 분석하는 방법으로는 Bag-of-Words방식을 이용한다. 이 방식은 클래스를 분류하여 학습시키고 테스트 데이터에 대한 분석을 통하여 각 클래스에 매칭시키는 기술이다. 또한 배경에 따른 여러 가지 불필요한 정보들이 존재하므로 이것을 최소화 하기 위하여 사진을 적절하게 분할하여 중요한 정보만을 추출하는 전처리 과정을 거친다.

본 논문의 구성은 다음과 같다. 2장에서 결과로 분류될 혼잡도의 개념을 정의 하며, 불필요한 정보를 최소화 하는 전처리 과정과 영역의 구분에 대한

기본적인 개념들을 설명한다. 3장에서는 Bag-of-Words와 사진에서의 응용을 기술하며, 4장에서는 실험 결과와 그에 대한 평가를 하고, 마지막 5장에서 문제점을 검토하고 연구방향을 제시하도록 한다. 논문에 쓰이는 각 사진들은 실험을 위하여 직접 차량에서 전방을 촬영한 결과이다.

### 2. 혼잡도의 정의 및 전처리 과정

혼잡도를 정의하는 일은 매우 어렵다. 그 이유는 사람마다 혼잡의 정도를 판단하는 기준이 다르고, 국가와 문화에 따라 교통혼잡에 대한 인식이 다르기 때문이다. 주관적인 가치를 객관화 시키기 위해서는 여러 사람이 그림에 대하여 평가를 해서 정의하는 방법이 있지만, 본 논문에서는 시간상의 문제로 위의 방법을 사용하지 못했다. 혼잡도 정의의 문제는 앞으로 계속 연구할 과제로 남겨둔다.



(a) 약간 혼잡

(b) 매우 혼잡

<그림 1> 차량 전방사진

본 논문에서는 혼잡도를 크게 3단계로 분류한다. <그림 1-b>와 같이 차가 많고 서 있는 상황을 ‘매우 혼잡’ 으로 설정하고, <그림 1-a>와 같이 전방에 차들이 있지만 달리고 있는 상황이나 전방의 차가 많지

않은 상황을 ‘약간 혼잡’ 으로 설정했다. <그림 2>와 같이 전방에 차량이 없거나 약간의 차량만이 원활하게 달리고 있는 경우 ‘원활’로 설정한다.

사진영상 하나만으로 혼잡도를 판단하기 위해서는 영상내부의 매우 많은 특징들 중에서 중요한 부분만을 추려내는 작업이 필수적이다. 그렇지 않으면 의미 없는 부분들이 많이 검출되어 잘못된 결과를 도출할 수도 있기 때문이다.



<그림 2> 차량 전방을 촬영한 사진. 원활한 도로 모습 전면에 빌딩과 아파트 등이 나타나 도로 혼잡도 판단에 관계없는 부분들을 다수 포함하고 있다.

<그림 2>에서 볼 수 있듯이 사진 안에 도로가 차지하는 영역은 적은 편이다. 그렇기 때문에 사진을 자동으로 필요한 부분 위주로 분할하여 분석하는 것이 좋은 결과를 도출할 수 있다.

전처리 과정에서 어떤 영역을 기준으로 분할해야 하는 것도 중요한 문제가 된다. 너무 많은 영역을 자르면 도로 전방의 중요한 데이터가 손실될 수도 있고, 너무 적은 데이터를 자르게 되면 앞에서 말한 것과 같이 불필요한 정보가 많이 남아있게 되기 때문에 결과에 영향을 줄 수가 있다. 그래서 본 논문에서는 적절한 절단 기준을 찾기 위해 500장의 사진을 무작위로 선정하여 전체적인 데이터의 질감을 표현하였다.



<그림 3>

차량의 전방에서 찍은 사진 500장 gray영상의 픽셀 값을 평균 낸 사진. 질감이 밑의 쪽에 많이 분포되어 있는 것을 확인할 수 있다. 오른쪽이나 왼쪽의 가장자리는 건물이나 반대편차선의 차량 등의 쓸모 없는 정보들이 존재한다.

<그림 3>에서 나타나듯 차량의 전방사진은 사진 하단 50% 정도에 중요한 데이터가 몰려있는 것을 확인할 수 있다. 그러므로 영상 분석에 들어가는 데이터는 모두

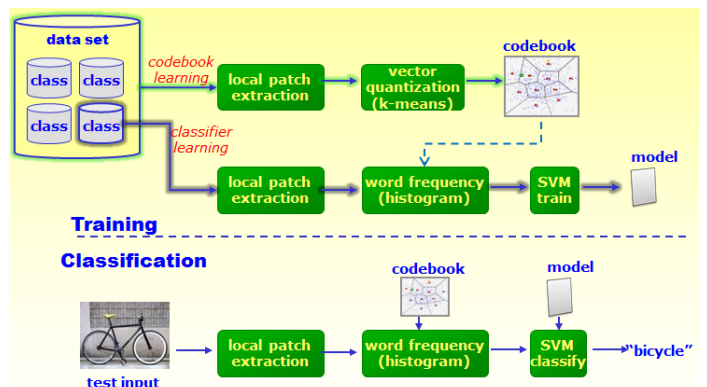
하단 50%만 들어가면 비교적 정확한 결과를 얻을 수 있다.

### 3. Bag-of-Words를 이용한 데이터 분석

Bag-of-Words는 원래 자연어 처리에서 텍스트를 분류하는 방법이다. 하지만 최근에는 Bag-of-Visual-Words라는 이름으로 영상 분석에 많이 사용되는 추세이다.[3] 본 논문에서도 Bag-of-Visual-Words를 이용하여 영상을 분석한다.

Bag-of-Visual-Words를 통한 데이터 분석은 이미지의 특징을 추출하여 벡터 양자화(vector quantization)를 통해 ‘Visual word’로 표현 되어 특정한 값으로 나타나는 방식이다. 이렇게 이미지를 벡터 값으로 표현한 후 이를 알고리즘을 통해 클러스터링 하여 시각적 단어들의 군집으로 표현한다. 각각의 클러스터들은 하나의 이미지 단어를 대표하고, 이 단어들의 집합이 단어집(vocabulary codebook)으로 구성된다. 단어의 개수를 얼마로 할 것인지 파라미터(parameter)에 따라 ‘codebook’의 사이즈가 결정된다.

위의 방식을 통해 각각의 이미지들이 특징들로 분류되어 고정된 차원의 벡터 양자화된 값으로 표현될 수 있다. 각각의 이미지들은 시각적 단어의 누적 개수에 따라 히스토그램과 같은 형태로 양자화 될 수 있고, 이렇게 양자화 된 이미지를 ‘매우 혼잡’과 같은 혼잡도 분류의 결론으로 도출할 수 있도록 하기 위해서는 분류된 카테고리 별 학습이 필요하다. 클래스 별 학습은 따로 사람이 태그를 붙여 분류해 놓은 이미지들을 히스토그램으로 만들어서 학습시키고 학습 모델을 추출해 내도록 한다.



<그림 4> Bag-of-Words

이렇게 학습된 모델을 이용하여 전방의 정보를 담은 입력사진이 들어가면 사진 내부의 특징들이 히스토그램화 되어 분류기에 의해 결론을 도출하게 된다. 이런 일련의 과정들이 어떤 알고리즘과 기술들을 통하여 이루어지는지 아래에서 자세하게 알아보도록

한다.

Bag-of-Words를 이용하기 위하여 먼저 local patch들을 추출해야 한다. Local patch들은 사진내부의 특징들을 의미하는데, 특징들을 추출하기 위하여 샘플링(sampling)과정을 거치게 된다. 샘플링에는 정밀 샘플링(dense sampling), 중요지점 샘플링(interest point sampling), 무작위 샘플링(random sampling)이 있다. 본 논문은 그 중에서 정밀 샘플링을 사용한다. 본 논문에서 제안하는 방식은 한 장의 데이터를 이용하여 교통상황을 파악할 수 있어야 하기 때문에 사진의 샘플링을 가장 정밀하게 해주는 방식이 적합하기 때문이다.

그리고 특징을 추출해 내는 방식으로는 SIFT(Scale Invariant Feature Transform)를 이용한다. SIFT알고리즘은 간단하게 특정 물체에서 그 사물을 가장 잘 표현할 수 있는 정보를 추출해주는 알고리즘이다. 이미지 내의 키 포인트를 추출하여 다른 이미지의 다른 각도나 크기의 같은 물체를 찾아낼 수 있도록 해주는 알고리즘이다. 그래서 이 알고리즘은 이미지의 크기와 회전에 강건한 특징을 추출할 수 있는 것으로 알려져 있다.[4] 본 논문에서 제안하는 방식은 질감을 이용하여 혼잡도 여부를 판단하기 때문에 전방의 차들에 대한 특징을 가장 잘 추출해 줄 수 있는 SIFT알고리즘이 적합하다. 본 실험에서는 32차원의 SIFT를 이용하도록 한다.

다음으로 전체적인 데이터들에 대한 단어(vocabulary)가 생성되어야 한다. 이 과정에서 많은 이미지 데이터들의 특징을 추출하여 그것의 대표 값을 선정하여 일반화를 시켜주는 과정을 거친다. 이때의 대표 값이 하나의 이미지 단어(image word)가 되는데, 이 단계에서는 K-means알고리즘을 사용하여 군집화를 하였다. K-means알고리즘은 간단하면서도 성능이 좋다는 장점이 있기 때문에 널리 사용되고 있다. 본 논문의 제안하는 방식에도 전체 시스템의 성능을 고려하여 K-means알고리즘을 이용하여 클러스터링을 하게 되었다. 이 단계를 거치면서 단어집(codebook)이 생성된다.

단어집 생성시 그 크기의 결정에도 이슈가 존재한다. 만약, 단어집의 크기가 작아지면 양자화 에러가 증가하며, 반대로 사이즈가 너무 커지면 차원이 증가하여 차원의 저주가 발생할 수 있다. 그러므로 일반화와 분별력간의 상충관계(trade-off)를 생각하여 단어집 사이즈를 적절하게 정해야 한다. 본 실험에서는 1K부터 4K까지 다양한 크기의 단어집을 생성하여 진행하였다.

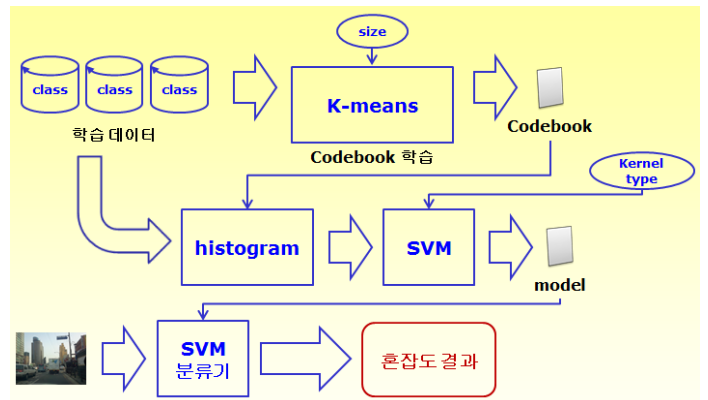
이미지 데이터에 대한 학습이 끝나면 이제 분류한 클래스 단위로 학습을 하게 된다. 본 논문에서 정의한 세 클래스 별로 데이터를 각각 입력하여 미리 생성된 단어집에 기반한 이미지 단어 빈도(word frequency)를 히스토그램을 통하여 확인한다. 본 논문에서는 사람이

느끼는 어느 정도의 혼잡을 통하여 미리 세 단계로 나눠놓은 각 350장씩의 이미지 데이터로 학습을 시켰다. 서론에서 언급한 것과 같이 혼잡이라는 개념은 여러 가지 조건에 따라 느끼는 정도가 달라질 수 있으므로, 다양한 방식으로 혼잡도를 객관적, 구체적으로 나타낼 수 있는 것은 앞으로의 연구 과제가 될 것이다.

그 후 SVM(Support Vector Machine)을 이용하여 학습을 시키고 학습 모델을 얻게 된다. SVM은 현재 알려져 있는 많은 방법 중에서 가장 인식 성능이 뛰어난 학습 모델의 하나이다. SVM의 경우 다른 알고리즘과 달리 최대 마진(margin) 점을 찾아서 분류를 하기 때문에 더 좋은 성능을 보이고 있다.[5] 본 실험에서는 libSVM을 이용하여 구현하였다.

위의 과정들을 통해 이제 실제 이미지 데이터를 입력하여 그 이미지의 특징을 추출한 후, 단어집에 맞춘 히스토그램을 얻고, 그 값을 다시 SVM에 넣어 미리 얻은 학습 모델을 기반으로 원하는 결과를 얻을 수 있게 된다.

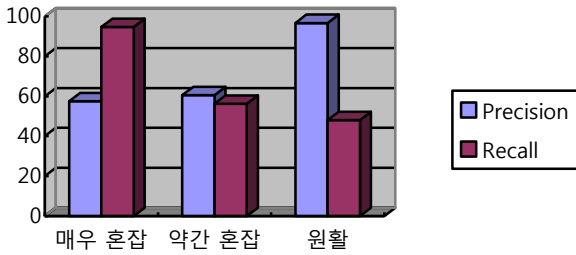
#### 4. 실험 결과 및 평가



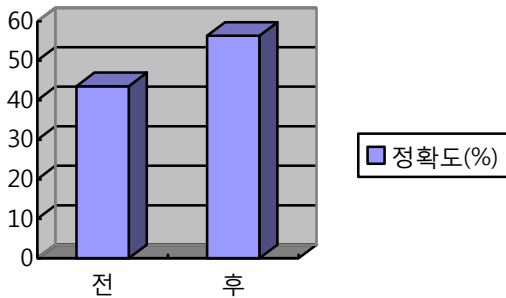
<그림 5>

본 논문의 Bag-of-words 적용의 예

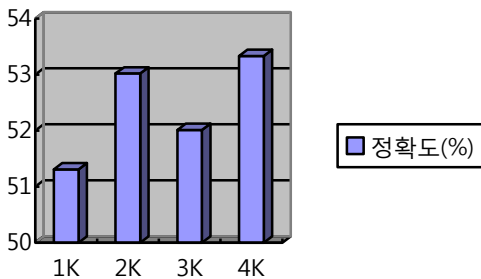
본 논문의 주요 아이디어는 차량 전방의 사진을 분석하여 질감으로 나타난 이미지를 세가지 클래스로 분류하여 정보를 전해주는 것이다. 그렇기 때문에 실험에 필요한 부분으로 나누는 전처리 과정으로 사진의 하반부 50%만 사용하도록 한다. <그림 5>은 본 논문의 실험을 간단하게 보여주고 있다. 실험은 Bag-of-Words를 이용하여 진행된다. 먼저 실험 이미지와 겹치지 않는 320\*240 크기의 이미지 600장에 대한 단어집 학습을 한다. 그리고 따로 세 클래스에 맞는 이미지를 각각 350장씩 카테고리화 시켜 각각의 학습모델을 만들도록 하였다. 그 모델들을 이용하여 실험 이미지 1,290장(각 클래스당 430장씩)을 입력하여 결과를 얻어냈다.



<그림 6> Bag-of-Words이용의 precision/recall



<그림 7> 사진의 전처리 전과 후



<그림 8> Codebook의 사이즈에 따른 전체 정확도

<그림 6>에서는 본 실험의 precision과 recall을 보여주고 있으며 <그림 7>에서는 전처리 과정을 거친 후 정확도가 더 향상되었음을 보이고 있다. <그림 8>은 codebook의 크기에 따라 정확도가 어떻게 나오는지 보여준다. 테스트 이미지에 대한 실험 결과가 매우 혼잡으로 많이 나타나기 때문에 위와 같은 결과를 보이고 있다. 이것은 이미지의 질감을 표현할 때 전방에 차가 없어도 도로의 질감이 마치 차와 같은 모습을 보이고 있어서 정확도가 낮아진 것으로 보인다. 대체로 ‘원활’의 사진을 넣었을 때는 ‘매우 혼잡’이나 ‘약간 혼잡’의 결과가 많이 나왔다. ‘약간 혼잡’의 경우의 사진을 넣었을 때도 ‘매우 혼잡’의 결과가 많이 나왔다. 하지만 ‘매우 혼잡’의 경우 매우 높은 확률로 정확하게 나오는 것을 확인했다. 이것을 종합해 볼 때, 클래스의 학습 과정에서 차량의

모습과 도로의 모습의 차이를 확실히 하지 못하여 생기는 문제로 파악 된다. 본 논문에서 제안하는 방법이 특정한 특징을 이용하여 비슷한 이미지를 찾는 것이 아니기 때문에, 도로상에 항상 존재하는 불필요한 정보들이 클래스 마다 학습되어 모든 클래스가 같은 사진으로 보이는 효과를 나타낼 수 있다고 생각한다. 좀 더 중요한 정보만을 표현하는 전처리 과정이 필요하다고 생각된다.

### 5. 문제점 검토 및 연구방향 제시

본 논문에서는 차량의 전방사진을 분석하여 사진의 질감을 통해 도로상황을 자동으로 판단할 수 있는 아이디어를 소개하였다. 차량 전방사진이라는 제한적인 요소에서 혼잡도라는 주관적인 결론을 내리는 것은 많은 제약이 있었다. 우선, 차량의 전방 사진이라는 특성상 수 많은 쓸모 없는 정보들이 포함되어 있었다. 혼잡도에 대한 정의를 객관화 시키기 위해 더 연구를 해서 근거를 찾아야 하고, 여러 가지 불필요한 정보가 많은 전방사진의 특징을 혼잡도의 특성에 맞게 분석하여 정확한 결론을 도출 할 수 있도록 하는 것이 앞으로의 과제가 될 것이다. Bag-of-Words방식의 문제보다는 질감을 이용하여 전방의 차량을 검출하는 문제가 실험의 정확도를 높이지 못한 요인이라고 판단한다. 전방의 차량이 붙어있는 상황이 많이 나타나서 도로와의 질감 차이를 나타내기 힘들고, 반대편의 차량 또한 불필요한 정보로 검출되지 못하고 차량의 일부로 포함되는 문제가 있었다. 앞으로는 전방 사진에 대한 더 정밀한 분석을 통해서 필요한 정보만을 추출하는 문제를 생각해 봐야 할 것이다.

### 참고논문

[1] 박종관, 강문희, “도시교통정책의 문제와 개선방향,” *한국행정학회 발표논문집*, 춘계학술대회, 2002.

[2] 최병걸, 정성일, 안철웅, 김승호, “Design and Implementation of a Road Congestion Analysis System using Regional Information,” *정보과학회 논문지*, 제5권, 제6호, 1999.

[3] J. Yang, C-W. Ngo, A. Hauptmann, and Y-G. Jiang, “Evaluation Bag-of-Visual-Words Representations in Scene Classification,” *In Proc. Of the ACM Multimedia Information Retrieval Workshop at ACM Multimedia*, pp 197-206, 2007.

[4] D-G Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” *International journal of computer vision*, Vol 60, pp 91-110, 2004.

[5] 석경하, 류태욱, “The Efficiency of Boosting on SVM,” *한국데이터정보과학회*, 제13권, 제2호, pp 55-64, 2002.