

A Study on Extracting Car License Plate Numbers Using Image Segmentation Patterns

Eun-Gyeom Jang*

Abstract

This paper proposes a method of detecting the license plates of vehicles. The proposed technology applicable to different formats of license plates detects the numbers by standardizing the images at edge points. Specifically, in accordance with the format of each license plate, the technology captures the image in the character segment, and compares it against the sample model to derive their similarity and identify the numbers. Characters with high similarities are used to form a group of candidates and to extract the final characters.

Analyzing the experimental results found the similarity of the extracted characters exceeded 90%, whereas that of less identifiable numbers was markedly lower. Still, the accuracy of the extracted characters with the highest similarity was over 80%. The proposed technology is applicable to extracting the character patterns of certain formats in diverse and useful ways.

▶ Keyword: Pattern, License plate, Text extraction, Pattern analysis

I. Introduction

2017년 12월을 기준으로 자동차 등록대수는 약 2,253만대로 인구 2.3명당 자동차 1대를 보유하고 있는 것으로 나타났다. 국토교통부는 2017년 12월 현재 우리나라 누적 등록대수가 지난해에 비해 72만 5,000대가 늘어났다고 집계되었다고 밝혔다[1].

자동차 대수의 증가는 교통체증, 교통사고 증가, 대기오염 증가, 주차 공간 부족, 소음 공해, 에너지 소비 증가, 산성비 등의 많은 문제를 발생시킨다. 이러한 사회적 문제가 있지만 자동차는 현대 사회에 반드시 필요한 필수품으로서 효율적인 활용 방안을 모색하여 문제를 해결해야하는 것이 우리의 과제이다.

자동차의 증가로 우리 실생활에 직면하는 문제점은 주차 문제, 자동차를 활용한 범죄 등을 들 수 있다. 현재 아파트, 학교, 공공장소 등에서는 효율적인 주차 관리를 위해 주차장 입출입시에 자동차 번호판을 관독하여 자동으로 요금을 계산하는 시스템이 구축되어 운영되고 있고, 또한 자동차 번호 판을 식별하여 교통 법규 위반 차량 관리, 뺑소니 자동차 및 범죄에 활용된 자동차 추적 등의 영역에 번호판 식별 기술이 적용되어 활용되고 있다[1,2].

그러나 자동차 번호판의 훼손 및 이물질이 묻은 번호판의 경우에는 번호 추출에 많은 어려움이 있어 명확한 자동차 번호가 추출되지 못하였을 경우에는 번호 인식 시스템에서 관독을 하지 못한다. 이러한 번호판 인식률을 높이기 위해 본 논문에서는 번호판 이미지를 분할하고 패턴을 분석하여 차량 번호 식별할 수 있도록 하였다. 논문 구성은 2장에서 번호판 추출 기술에 대한 관련연구를 분석하고, 분석한 기존 기술에 문제점 및 효율적 접근을 위한 제안 기술을 3장에서 기술한다. 4장에서는 제안한 기술을 실험하여 성능을 분석하고 마지막 5장에서는 연구에 대한 결론으로 구성하였다.

II. Related research

2.1 Number plate analysis

1) Number plate Type

차량 번호판은 모든 자동차의 전·후면에 직사각형으로 부착

*First Author: Eun-Gyeom Jang, Corresponding Author: Eun-Gyeom Jang
*Eun-Gyeom Jang (jangeeg@jangan.ac.kr), Dept. of Internet Communication, Jangan University
• Received: 2018. 08. 01, Revised: 2018. 09. 17, Accepted: 2018. 09. 27.
• The Work was supported by Jangan University Research Grant in 2018

한다. 자동차등록 번호판은 한글과 아라비아 숫자의 조합으로 이루어진 일련번호이다. 해가 갈수록 자동차 수의 급격한 증가로 색상 및 유형이 바뀌었다. 1973년부터 “자동차의 사용목적(자가용/사업용) + 한글 한 글자 + 일련 번호(1000~9999)”으로 구성된다. 1973년 4월, 1996년 1월, 2004년 1월 개정되어 현재 부분적인 추가 사항 및 변경으로 사용되고 있다. 기본 형식은 “00 가 0000”으로 승용차(01~69), 승합차(70~79), 화물차(80~97), 특수차(98~99)로 용도에 따라 구분하고, 용도기호(가~마, 거~저, 고~조, 구~주)로 32개로 구분하여 사용한다(2013년 자가용 자동차에 “하”, “호” 새롭게 추가). 자동차등록 번호판 유형은 그림 1과 같다[1,2,3].

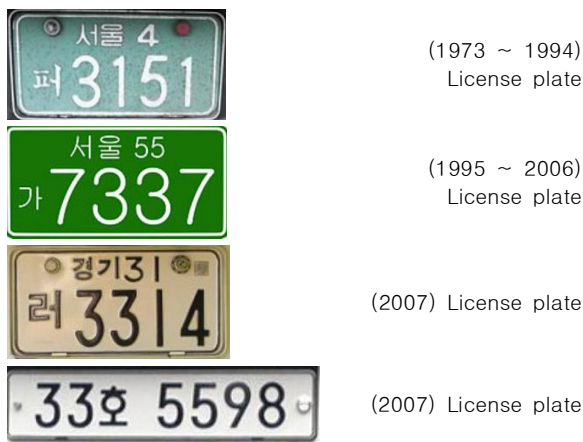


Fig. 1. Number plate Type

2005년 2월부터 경찰차와 관용차에 시범으로 적용하던 유럽형 번호판(일자형 흰바탕 검은글씨)을 글씨체와 용도기호의 크기를 변경하여 2006년 11월부터 모든 자가용 자동차에 시행하였다[1,4,5]

2) Number plate type analysis

차량 번호판은 1973년을 기준으로 다양한 형태를 갖는다. 용도에 따른 구성과 형식, 색상, 개정시 마다 변하는 색상과 구조들을 볼 수 있다[1,6]논문에서는 유럽형 번호판(그림2)을 기본 모형으로 분석하여 제안 기술을 적용한다. 번호 추출을 위한 기본 알고리즘 측면에서 번호판을 분석한다.

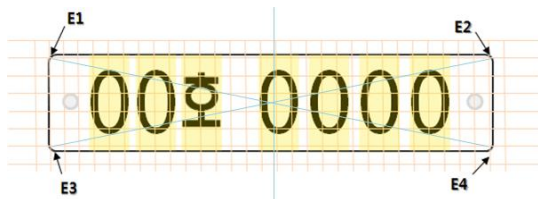


Fig. 2. Number plate format

유럽형 번호판은 가로(≒0.84), 세로(≒0.16)의 비율로 구성된다. 번호판은 앞 숫자 두 자리, 한글 한 자리, 뒤 숫자 네 자

리, 모두 일곱 자리의 숫자와 글자가 가로로 한 줄로 구성된다.

E1을 중심으로 E2, E3, E4의 거리 비율, E1과 E4, E2와 E3의 거리 비율로 이미지의 기울기를 구할 수 있다. 이미지의 기울기 비율에 의해 번호판의 정형적인 규격의 정형화를 통해 복원된 영역별 이미지 패턴을 추출할 수 있다. 또한, 4개의 에지 포인트가 아닌, 번호판의 정형화된 숫자 및 글자의 배치 패턴에 의해 번호판이 규격화되어 있다. 이러한 번호판의 특성을 고려하여 규격화된 번호판의 이미지 패턴을 통해 차량의 번호를 추출할 수 있도록 블록 패턴 방식을 적용하여 숫자 및 글자를 인식할 수 있는 기법을 제안한다.

2.3 Image extraction

1) Extracting particular segments using color information

흑백 명암의 변화를 측정하여 영역을 분할하는 기술은 빛에 강한 특징을 가지고 있어 빛에 대한 정보 손실이 적고 빠르게 처리할 수 있는 특징이 있다. 하지만 양극화된 정보로서 형태가 비슷한 경우에는 차별성이 있게 값을 추출하지 못하는 문제가 발생하기도 한다. 또한 명암의 변화에 따른 번호 추출 기법은 임계값에 영향을 많이 받는다. 적절한 임계값을 적용하기 위해 이미지 처리 시간과 임계값 조정에 따른 속도 저하가 발생하고 있다. RGB와 HSI 컬러의 특징을 활용한 기법은 주의의 밝기 값에 영향을 받아 환경적인 요인에 매우 민감하게 반응하고 있다. 이러한 주변의 환경에 민감함을 보완하기 위해 HSI의 H 성분과 Q 성분을 결합하는 기술을 적용한다. 번호판 영역 추출 과정은 다음과 같다[7].

- ① 원영상의 RGB 값을 추출한다.
- ② 일정한 간격으로 RGB 값을 HSI와 YIQ로 변환
- ③ H 성분과 Q 성분을 추출
- ④ 각 성분을 0 ~ 255구간으로 정규화
- ⑤ 추출 데이터 이진화
- ⑥ 이진 데이터의 수평 프로젝션 적용
- ⑦ 적응적인 임계값을 사용하여 라인 결합하여 세로 후보 구간들을 추출
- ⑧ 수직 프로젝션 적용 후 가로 후보 구간들을 추출하여 번호 영역을 추출

에지투영 및 방향성에서는 번호판의 수평영역으로 구분하여 수직영역으로 번호판 영역을 추출한다. 수평영역 추출은 입력 영상의 평균 밝기 값을 계산하고 평균 밝기 값을 고정시켜서 번호판 인식률을 높였다. 수평방향의 투영데이터를 구하기 위해 마스크 패턴을 활용하였다.

2) Comparison of pixel units and histograms

동일한 샷 내에서는 화소 값의 변화가 적은 특성으로 i번째 프레임에서 (x, y)의 화소 값의 차이를 비교하여 임계값보다 크

면, 다른 영역으로 인지하여 이미지 영역을 추출한다. 특정 임계값을 넘으면 다른 영역으로 구분하여 처리한다. 화소 값은 식 1과 같다[7].

$$\frac{\sum_{x,y=1}^{X,Y} (\text{if } |F_i(x,y) - F_{i+1}(x,y)| > t)}{X*Y} * 100 > \text{임계값} \quad (1)$$

이러한 화소단위 비교법은 이미지의 정형적인 특성과 임계값에 의해 이미지 비교의 결과가 판이하게 다른 결과를 추출하기도 한다. 즉, 페이드와 줌 인/아웃과 같은 영상의 변화, 물체 이동은 많은 화소의 변화가 있어서 잘못된 검출결과를 갖는다.

히스토그램 비교법은 유사한 이미지 색상분포에 대한 보편적인 검출방법이다. 동영상의 경우 인접한 프레임들을 히스토그램 차이값으로 특정한 임계값으로 장면의 차이를 추출한다. I와 j를 각각 프레임번호, 히스토그램상의 색상 값, H를 색상 값에 대한 빈도라 할 때 다음 식에 의해 이미지의 차이를 구할 수 있다(식2).

$$D_i = \sum_{j=1}^N |H_i(j) - H_{i+1}(j)| T_h \quad (2)$$

히스토그램 방식은 동영상의 경우, 하나의 컷에서는 각 프레임이 비슷한 특성을 유지하여 이미지의 차이가 적을 때 유리한 특성을 갖는다. 그러나 히스토그램 비교법은 밝기의 변화에 예민한 특성을 갖는다. 조명과 빛에 예민하게 반응하여 잘 못된 이미지 비교 결과를 얻을 수 있는 단점을 갖는다.

3) Comparison of entropy and edge images

엔트로피는 확률을 이용한 기법으로 두 이미지의 엔트로피 차이로 영상을 비교한다. 같은 영상이라 할지라도 조명의 변화가 발생하면 히스토그램 방법은 잘못된 영상 비교 값을 추출하나, 엔트로피 값은 밝기 변화가 있어도 영상의 복잡도에 영향이 적다. 히스토그램의 비교법과 엔트로피 비교법을 같이 활용하여 어느 특정한 임계값으로 이미지를 비교하여 영상의 유사도를 판단한다. 히스토그램 값이 H_{Value} 이고, w는 가중치, 엔트로피 값이 E_{Value} 일 때 다음과 같이 임계값 I_{Value} 은 식3으로 구할 수 있다.

$$H_{Value} + w(E_{Value}) > I_{Value} \quad (3)$$

특정 임계값을 기준으로 이미지의 유사도를 판단할 수 있다. 그러나 엔트로피는 밝기 변화에 강력한 장점을 갖지만, fade 서서히 밝기의 변화에는 잘 못된 영상 비교 결과를 추출하는 단점을 갖는다.

압축영역에서의 에지 영상 비교법은 AC 계수에서 저주파 5개의 강도 PL과 임계값으로 에지 블록을 계산하여 에지 영상을 구한다. 참조 프레임의 수평과 수직의 에지 히스토그램과 현재

프레임의 에지 히스토그램을 비교하여 영상의 차이 값을 추출한다(식4, 식5).

$$P_L = F_{01}^2 + F_{10}^2 + F_{02}^2 + F_{20}^2 + F_{11}^2 \quad (4)$$

where, $F_{i,j}$: (i,j) 위치의 DCT계수

$$d(k, k+1) = \sum_{x=0}^{M-1} |P_H(x+u) - P_H(x)| + \sum_{y=0}^{N-1} |P_V(y+v) - P_V(y)| \quad (5)$$

where, M, N : 에지 영상의 가로, 세로 크기
 u, v : 수평, 수직 방향의 움직임 벡터

AC 계수를 이용한 것 보다 DCT 계수를 활용하는 것이 이미지 비교에 정확한 결과를 얻을 수 있다. 밝기 값이 변하는 블록에는 DCT의 DC 크기를 디코딩하여 이미지 비교 처리 시간이 빠르다는 장점을 갖는다.

III. Text extraction technique using block pattern

3.1 Process for extraction block patterns

그림 3은 번호판 이미지에서 숫자 및 한글 텍스트를 추출하는 과정이다.

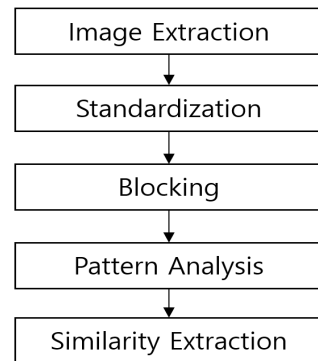


Fig. 3 Pattern Extraction Process

- ① Image Extraction : 카메라로 찍은 차량의 번호판 영상 이미지를 추출
- ② Standardization : 추출한 이미지를 번호판 규격에 맞도록 크기 비율을 적용하여 표준 모델로 변환
- ③ Blocking : 표준화된 이미지를 한글 텍스트 및 숫자 영역으로 구분하여 글자별로 블록을 나누고 블록별 패턴 처리를 위해 세분화 블록 생성
- ④ Pattern Analysis : 세분화된 블록의 패턴 영상을 화소 및 이미지 처리 기법을 활용하여 분석

- ⑤ Similarity Extraction : 분석된 이미지 패턴을 표본 모델과 비교하여 수치화

3.2 Image Standardization

유립형 번호판의 규격은 가로, 세로 비율이 0.84, 0.16을 갖는다. 번호판을 보는 각도에 따라 각 에지 값을 기준으로 비율이 달라진다. 그림 4와 같이 보는 방향에 따라 에지의 기울기가 다른 것을 볼 수 있다.

Image	Edge Slope

Fig. 4. Edge slope of Image

에지의 기울기는 4개의 에지 라인을 통해 기울기 정도를 알 수 있다. 다음 그림 5와 같이 기울기 및 에지 점을 표현할 수 있다.

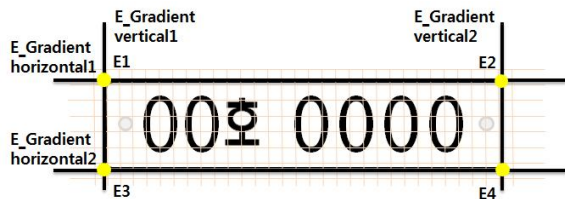


Fig. 5. Image Edge Point

$E_Gradient\ horizontal1$ 과 $E_Gradient\ horizontal2$ 는 가

로방향의 포인트로서 가로방향의 기울기이고 $E_Gradient\ vertical1$ 과 $E_Gradient\ vertical2$ 는 세로방향의 기울기를 나타낸다. $E_Gradient$ 은 $E_{(1-4)}$ 을 기준으로 기울기 값을 얻을 수 있다. E1의 에지 포인트 $E1(x, y)$ 를 중심으로 E2의 에지 포인트의 위치값 $E2(x,y)$ 로 $E_Gradient\ horizontal1$ 의 기울기를 얻는다. $E_Gradient\ horizontal2$ 또한 $E3(x, y)$ 와 $E4(x, y)$ 의 위치 값으로 기울기를 얻을 수 있다. 기울기 값은 식 6으로 구할 수 있다.

$$E_Gradient = \frac{E_{i+1}y_{i+1} - E_i y_i}{E_{i+1}x_{i+1} - E_i x_i} \quad (6)$$

(The slope value between E_i and E_{i+1})

기울어진 이미지는 가로, 세로 방향으로 정규화하여 정확한 번호판의 규격 비율로 복원하여 처리한다. E1의 (x, y) 값을 중심으로 $E3(x, y)$ 의 가상 위치를 선정하고 선정된 값을 중심으로 E2와 E4의 위치 값을 조정한다. E2의 위치 값은 규격화된 번호판의 크기 비율로 적용하여 4개의 에지 포인트를 가상 포인트로 선정하여 현재의 에지 포인트와 가상 위치 포인트를 거리 및 이동 벡터를 구하여 이미지를 확장 및 축소하여 규격화한다. 이미지 규격화를 위한 절차는 그림6과 같다.

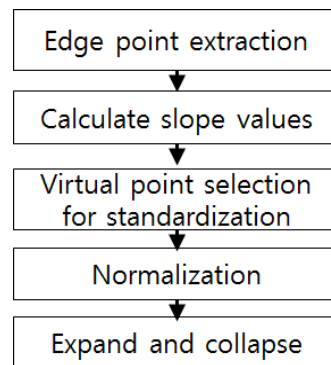
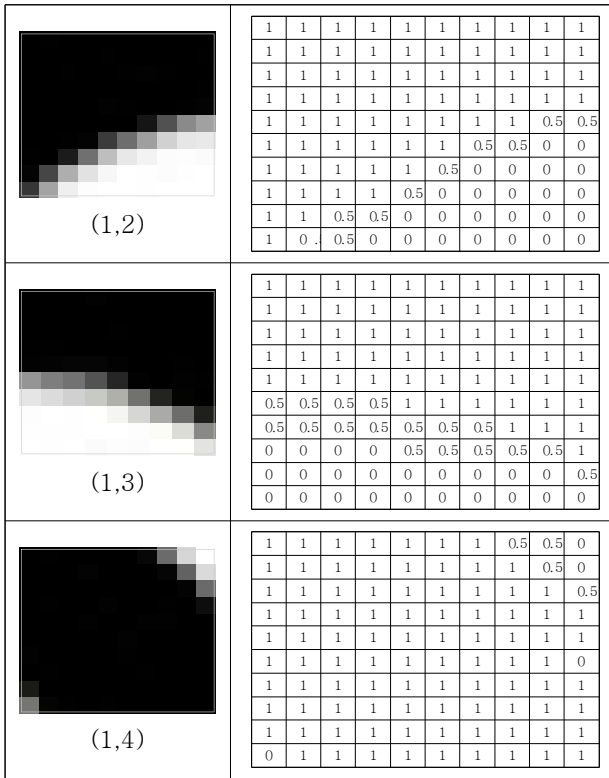


Fig. 6. Image Standardization Process

- ① 에지 포인트 추출 : 주어진 영상에서 E1, E2, E3, E4의 (x,y) 포인트를 추출하여 번호판 영역의 포인트를 도출
- ② 기울기 값 구하기 : 추출된 영상의 에지 포인트를 중심으로 $E_Gradient\ horizontal1$, $E_Gradient\ horizontal2$, $E_Gradient\ vertical1$, $E_Gradient\ vertical2$ 값 계산
- ③ 표준화를 위한 가상 포인트 선정 : 기울어진 영상을 규격화된 모델로 복원하기 위해 번호판 규격 비율에 맞는 가상 이미지 생성
- ④ 정규화 : 가상 규격 포인트에 E(1~4)의 에지 값을 계산하여 (x, y) 포인트를 이동
- ⑤ 확장 및 축소 : 정규화에 의해 이동한 에지 포인트를 중심으로 벡터 값을 구하여 전체 이미지 영역을 축소 및 확장



4.2 Analysis of Experiment

제안 모델의 성능을 분석하기 위해 번호판 이미지를 적용하여 테스트한다. 실험을 위한 번호판 표본은 개인정보보호를 위해 비공개 모델로 사이즈 및 촬영 각도의 변화, 손상 및 이물질에 의한 번호 식별의 어려운 상황으로 테스트 하였다.

촬영된 이미지는 이미지 표준화 과정으로 기울기가 조정되고 에지 포인트를 중심으로 영역별 글자 이미지를 추출한다. 추출한 이미지는 표준 모델과 비교하여 유사도 값을 구하고 최종 인식된 번호를 도출한다.

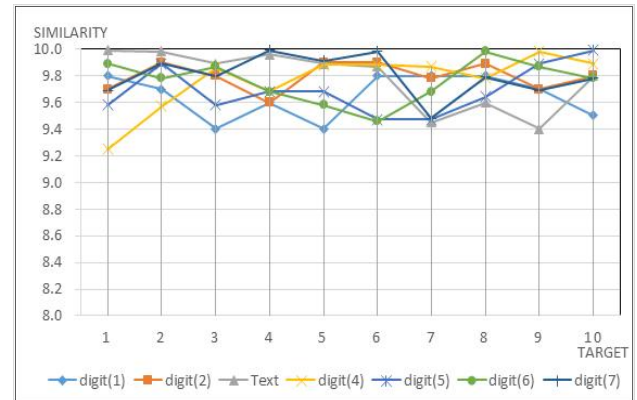


Fig. 11. Normal Image

그림 11은 일반적인 번호판 이미지를 촬영한 이미지로서 10개의 번호판을 제안기술로 유사도를 분석한 결과이다. 적게는 92%의 유사도를 갖는다.

IV. Efficiency Analysis

4.1 Block Pattern Analysis

번호판의 표본 모델과 비교 대상의 번호판은 3.3절에서 변환된 블록 양자화 값에 의해 패턴이 비교된다. 표본 모델을 비교 대상의 모델에 $f: (x,y) \rightarrow (x,y)'$ 의 관계로 값을 추출한다. 표본 모델의 행렬블록 (x, y) 는 식 8로 얻을 수 있다.

$$f(x,y) = f(x,y)' : true$$

$$f(x,y) \neq f(x,y)' : false$$

$$\begin{matrix} f(x,y) & f(x,y)' \\ \left(\begin{matrix} (x_1,y_1) \\ (x_2,y_2) \\ (x_3,y_3) \\ \vdots \\ (x_{n-1},y_{n-1}) \\ (x_n,y_n) \end{matrix} \right) & = \text{or} \neq \left(\begin{matrix} (x_1,y_1)' \\ (x_2,y_2)' \\ (x_3,y_3)' \\ \vdots \\ (x_{n-1},y_{n-1})' \\ (x_n,y_n)' \end{matrix} \right) \end{matrix} \quad (8)$$

비교대상의 모델은 허용번호의 표본 모델을 순서적으로 비교하여 유사도가 제일 많은 표본 모델을 추출한다. 이렇게 선정된 표본 모델들을 조합하여 숫자를 완성하고 이미지 전체 패턴으로 모델을 검증한다.

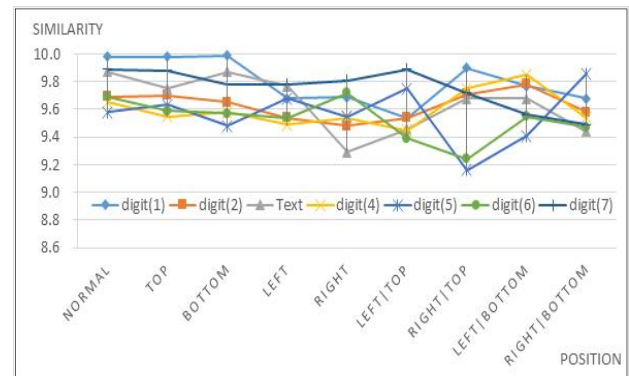


Fig. 12. Image by Location

그림 12는 분석하기 위한 번호판을 정상적으로 앞에서 촬영하고 위/아래/왼쪽/오른쪽/위/왼쪽 아래/오른쪽 위/오른쪽 아래에서 촬영한 이미지를 분석한 결과이다. 최소 약 91%의 유사도를 가지며 일반적으로 촬영한 이미지와 비교하여 낮은 유사성을 갖는 것으로 분석되었다.

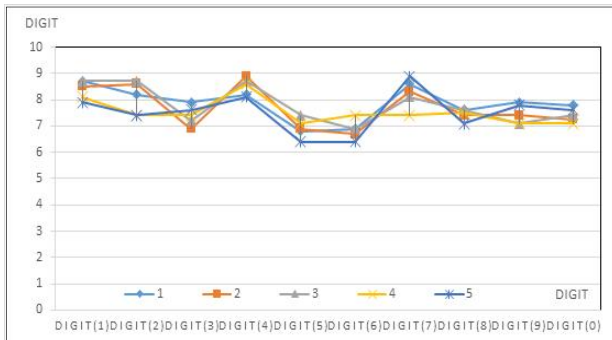


Fig. 13. Similarity of Number

그림 13은 각 번호별로 유사성을 측정하기 위해 실험한 결과이다. 각 숫자가 다른 숫자와의 차별적 유사성을 갖는지를 분석한 실험이다.

기존 명암벡터[8]를 위한 기술은 차량번호판의 기울기 및 노이즈에 의한 훼손된 영역 경우, 번호 인식이 낮은 반면 제안 기술은 보다 높은 성능을 발휘하였고, 번호판 이미지 기울기 조정을 통한 번호 추출기술[9] 보다 명암 대비 글씨 추출 기법을 활용하여 높은 인식 및 성능을 보였다.

제안 모델의 실험 결과 약 90% 이상의 실험 결과를 얻었다. 다소 낮은 유사성을 갖는 경우도 있지만, 각 숫자의 표준 모델과 각기 다른 숫자를 비교한 결과는 번호를 판별하는데 월등한 차이의 유사성 결과를 도출하여 차량 번호를 인식하는데 어려운 점은 없었다. 하지만, 번호판의 훼손이나 식별하기 어렵도록 이물질이 묻은 번호판에 대해서는 신뢰할 만한 데이터를 추출하지는 못하였다. 하지만 유사성을 비교한 결과로는 유추할 수 있는 데이터 분석결과를 얻었다.

V. Conclusion

본 논문에서는 자동차 번호판의 글자를 추출하기 위한 기술을 제안하였다. 번호판 이미지를 표준화하고 예지 포인트로 글자에 대한 이미지 영역을 추출하여 각 영역별 글자의 패턴을 분석하여 유사성을 도출하였다. 이렇게 도출된 값은 표준 모델과 비교하여 유사성이 제일 높은 글자를 추출하는 기술이다.

본 연구에 활용된 자동차 번호판 모델은 2007년에 개정된 차량용 번호판을 표본으로 적용한 기술이다. 2019년에 새롭게 자동차 번호판 모델을 개정하기 위해 추진 중에 있으며 포맷이 다른 과거의 자동차 번호판에 제안 기술을 직접 적용하기에는 어렵다. 그러나 본 논문에서 제안한 기술을 각 모델별로 예지 포인트로 영역별 글자를 추출하여 적용한다면 추후 개정된 차량 번호판을 포함한 글자 추출 기술을 요하는 영역에 유용하게 활용될 것으로 기대한다.

이 논문은 장안대학교 2018년도 연구비 지원에 의하여 연구되었음.

REFERENCES

- [1] Korea Insurance News, "Vehicle registration number 225,300,000, 2.3 per person" "http://www.insnews.co.kr/design_php/news_view.php?num=53184&firstsec=1&secondsec=14" 2018.1.21.
- [2] Young Min Jun, Jeong Hee Cha, "An Algorithm for Segmenting the License Plate Region of a Vehicle Using a Color Model", The Institute of Electronics Engineers of Korea - Computer and Information 43(2), March. 2006.
- [3] Seong-Woo Kim, Dong-Koo Kang, Jae-Hyun Park, Eui-Young Cha, "The Recognition of Vehicle Plate's Korean Character Using Grapheme Segmentation", Korea Information science society, 29(1B), April. 2002.
- [4] Young Sik Park, "A License Plate Extraction Using Wavelet Transform", Journal of Korea Institute of Information Technology, 6(2), April. 2008.
- [5] SungWon Mun, Undong Chang, YoungJun Song, JaeHyong Ahn, "A performance analysis of car license plate recognition to slants", Proceeding of KIIT Summer Conference, August. 2004.
- [6] Hyoung-Chul Oh, Jong-Ho Choi, "A Recognition Algorithm of Car License Plate using Edge Projection and Directivity Vector", Journal of Korean Institute of Information Technology, 7(1), February. 2009.
- [7] Eun-Gyeom Jang, "Authentication Mechanism for Protecting Digital Contents Right in Internet UCC Environment", Daejeon Univ. Graduate school, Doctoral thesis, August. 2007.
- [8] Sook Kim, Hyung Ki Cho, Joon Young Min, Jong Uk Choi, "A License Plate Extraction Algorithm using Brightness Vector", Journal of KISS(B): Software and Applications 25(4), pp. 676-684, April. 1998.
- [9] SungWon Mun, Undong Chang, YoungJun Song, JaeHyeong Ahn, "A performance analysis of car license plate recognition according to slants", Proceedings of KIIT Summer Conference, Korean Institute of Information Technology, pp 90-93, August. 2004.

Authors



Eun-Gyeom Jang received a Ph.D in Daejeon University in 2007. Hi is currently a Professor in the Department of Internet Communication Jangan University.

It has an interest in mobile communications, system security and Computer Forensics.