

## Program Assignment #2

- Due Date : Dec. 8 (Fri) 23:59
- Submission : 소스코드 & 보고서(표지 제외 10장 이하로 요약하여 작성)

### <프로그램 과제 개요 (필기체 글자 및 단어 인식)>

모든 패턴인식 문제에서, 높은 성능을 얻기 위해서는 적절한 특징 벡터를 사용해야 한다. 또한 적절한 전처리 과정을 통해 잡음, 주변 환경 등의 영향을 최소화 하는 과정도 필요로 한다.

본 과제는 필기체 글자 및 단어 인식 시스템을 디자인하는 것을 목표로 한다. 필기체 인식을 위한 전처리 과정, 특징벡터 추출 과정, 인식 과정을 설계하고, 이를 이용한 필기체 인식 프로그램 및 보고서를 제출하시오.

### <Task 요약>

#### 1) 전처리 (Preprocessing)

##### - Segmentation (글자, 단어)

- 주어진 파일을 개별적인 글자로 분리해 내는 것 (데이터 설명 참고)
- 단어의 경우, 단어를 글자 단위로 분리하는 과정까지 포함

##### - 회전 조절 (Optional)

회전된 글자를 일괄적으로 조절하는 작업

##### - 굵기 조절 (Optional)

글자마다 굵기가 다른 것을 일괄적으로 보정하는 작업

- 그 밖에 본인이 필요하다고 생각하는 전처리 과정이 있다면 설계하고 구현할 것, 위에서 언급된 내용들도 필요로 하지 않는다면 구현하지 않아도 무방

#### 2) 특징 추출 (Feature extraction)

- 각 알파벳을 구분하기 위한 본인만의 특징 벡터 추출 방법 제시 및 구현

#### 3) 인식기 (Classifier)

- Segmentation 이 이루어진 개별 글자, 및 단어에 대해 인식 수행
- 단어의 경우 글자열(sequence)을 모두 맞췄을 경우 정답으로 인정

### <참고사항>

- 하나의 통합된 시스템 상에서 알고리즘이 동작하는게 가장 바람직함 (가산점 부여 예정)
  - 즉, 인식기의 입력으로 글자가 들어오면 글자 단위의 인식을, 단어가 들어오면 단어 단위의 인식을 하도록 자동화된 시스템 구현
- 글자 인식기와 단어 인식기를 별개로 구현하는 것도 허용함

<데이터 설명>

1. Training Data: TRAINING\_DATA 폴더내의 이미지 파일.

필기체 a, b, c, d, e의 각 문자열 60개로 이루어진 파일을 gray scale 및 binary scale(흑백)로 제공.  
(원하는 스케일로 선택하여 사용)

2. Test Data: TEST\_DATA 폴더내의 이미지 파일 (gray scale)

Ex)



<보고서>

- 전처리 과정, 사용한 특징벡터, 사용한 Classifier, **Test Data를 이용한 인식률**, 프로그램 실행 방법들을 정리하여 보고서로 제출(주어진 양식은 없으나 위 항목들은 반드시 포함할 것)
- 표지제외 10장 제한.
- 이론적 내용은 간단하게 기술, 실제 구현 방법 위주로 작성.

<주의사항>

1. 개발 환경

사용언어 및 툴: MATLAB

2. 보고서

표지 제외 **10장으로 제한**. 넘을 경우 감점.

3. 제출

구현한 프로그램의 프로젝트 파일과 보고서를 압축하여 제출.

제출할 파일명: **이름\_학번.zip ( ex) 이승형\_2015xxxx.zip )**

4. 평가

조교가 보유 중인 있는 또 다른 Evaluation Data를 이용하여 평가할 예정이므로, overtuning 지양할 것.

문자인식률(60), 단어인식률(30), 보고서(10)

Good Luck!