

Motivación

El reconocimiento de patrones es una disciplina científica que incluye métodos para adquirir, procesar, analizar y comprender las imágenes del mundo real con el fin de producir información numérica o simbólica para que puedan ser tratados por un ordenador.

Es una de las áreas más exploradas entre las Ciencias de la Computación debido a su gran valor en aplicaciones prácticas, como la detección de tumores, reconocimiento facial en sistemas de seguridad, control de calidad en líneas continuas de producción, entre otras.

Exploraremos el ejercicio de detección de caracteres a partir de imágenes que contienen información de las letras A, D, F, G, S, utilizando técnicas de segmentación, extracción de características geométricas, normalización de datos, clasificación y regresión.



Figure 1: Procesamiento y segmentación de imagen original.

Solución propuesta

Para enfrentar este problema se decidió utilizar métodos de cálculo de contornos para la segmentación de imagen original. Para la extracción de características, obtuvimos features basados geometría, específicamente los Momentos de Hu, Roundness, área y perímetro.

Luego, para cada feature, se aplicó una estandarización Z-score. Por último, se presenta la función **predict()**¹, la cual esta basada en un clasificador KNN, entrenado con los 5 primeros Momentos de Hu y una cantidad unitaria de vecinos.

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma}, \quad \gamma = 1 + \frac{p+q}{2}$$

$$H_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$$

$$H_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$

$$H_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} - \eta_{03})^2$$

$$H_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$

Figure 2: 5 Momentos de Hu.

Experimentos

Se construyó un dataset constituido de las características y etiquetado de las imágenes. Para esto recorrimos el array de la imagen binaria, segmentando el array según la dispersión de sus píxeles. Luego, se procedió a binarizar cada segmento para extraer información basada en descriptores geométricos.

Por último, se implementó un algoritmo del tipo K-Nearest Neighbors, el cual computa las distancias para calcular qué tipo de dato podemos predecir a partir de sus características.

Se realizaron múltiples experimentos con el fin de obtener los hiperparámetros que maximizaran el accuracy de las predicciones. Para esto, se utilizó cada combinación de features geométricos y de número de vecinos, resultando finalmente en una elección de los 5 primeros Momentos de Hu como mejores features geométricos y una cantidad de K=1 Neighbors. Se calculó el accuracy para cada iteración desde K=1 hasta K=40.

El motivo de la elección de los 5 Momentos de Hu se debe a que estos fueron calculados con invariabilidad frente a tamaño, rotación y posición², lo cual los hace ideales para enfrentar este caso.

En los primeros experimentos presentaron resultados por bajo lo esperado, obteniendo 74.41 en accuracy con K=1 sin normalización. Una vez normalizados los datos, mejoramos nuestro rendimiento hasta obtener un máximo de 99.00.

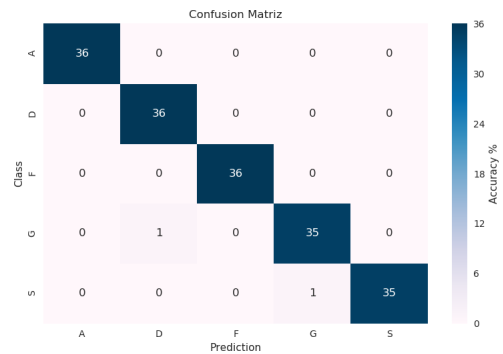


Figure 3: Matriz de confusión.

Conclusión

Luego de experimentar múltiples combinaciones de segmentación, extracción y uso de características y construcción del modelo predictivo, pudimos apreciar la incidencia del control exhaustivo de cada paso en nuestro resultado final.

Pudimos apreciar la incidencia de la correcta segmentación, normalización de datos y elección de features en los resultados. En un principio se utilizó un dataset con imágenes cortadas manualmente, lo que demostró influir negativamente.

En un mismo ejercicio, fue posible observar comportamientos muy por debajo de lo esperado debido a una falta en el control de la dispersión de nuestros datos y mal selección de features.

Podemos concluir de los experimentos realizados que estos descriptores geométricos son una buena aproximación inicial para enfrentar este problema. Sin embargo, es un método a medida y podría romperse al incluir otro tipo de letras o signos. Queda como trabajo futuro la ampliación del set de letras y signos a analizar, además de expandir el rango de técnicas y descriptores usados en nuestro modelo predictivo, con el fin de aumentar la robustez de nuestro sistema.

¹Contenida en el archivo characteristics.py

²M. K. Hu, "Visual Pattern Recognition by Moment Invariants"