

# Tarea de Investigación 6

1<sup>st</sup> Ericka Céspedes Moya  
Tecnológico de Costa Rica  
San José, Costa Rica  
ericka.cespedes@gmail.com

2<sup>nd</sup> Esteban Alonso González Matamoros  
Tecnológico de Costa Rica  
San José, Costa Rica  
esteb.gonza29@gmail.com

**Resumen**—El siguiente documento contiene la primera y segunda parte de la Tarea de Investigación 6 para el curso de Introducción al Reconocimiento de Patrones. Esta tarea consistió en utilizar un grupo de 5 especímenes que ya habían sido utilizados para la Tarea de Investigación 4 con el propósito extraer el vector de los Momentos Invariantes Hu y utilizarlo como reconocedor de vocales. Además, se analizan y comparan los resultados obtenidos de los tres reconocedores implementados.

**Index Terms**—procesamiento de imágenes, patrones, reconocimiento, Python, OpenCV, momentos, Hu, vocales, comparación

## I. 7 MOMENTOS INVARIANTES PARA CADA VOCAL

Primero que todo, es necesario mostrar los 5 especímenes utilizados para todas las pruebas de este informe, los 5 especímenes que son 5 vocales son los siguientes:



Figura 1. Vocal a utilizada



Figura 2. Vocal e utilizada

En el caso de la i, tuvimos un problema debido a que se recorta sin el punto de arriba de la vocal, sino que solo utiliza la línea de abajo.



Figura 3. Vocal i utilizada



Figura 4. Vocal o utilizada



Figura 5. Vocal u utilizada

Para esta comparación se utilizará una tabla donde se indican todas las vocales y los momentos de Hu respectivos para cada una de ellas.

Vocal	Hu 1	Hu 2	Hu 3	Hu 4	Hu 5	Hu 6	Hu 7
a	2.811	7.799	10.164	10.532	-20.922	14.435	21.259
e	2.967	7.919	10.328	10.777	21.382	15.159	21.665
i	3.189	8.335	12.299	14.120	27.381	-18.953	-27.672
o	2.856	7.488	11.734	11.636	-24.146	15.734	-23.327
u	2.978	7.462	10.745	10.807	-21.592	-15.399	22.276

## II. COMPROBACIÓN DE INVARIACIÓN EN MOMENTOS DE HU

Para comprobar la invariación de la rotación, desplazamiento y escalación, lo que decidimos realizar fueron 3 tablas donde cada una posee 4 variaciones y que mostrarán los momentos de Hu 1 y 7 de cada una de ellas, no se agarran los 7 momentos de Hu porque ocuparían mucho espacio en las tablas. En esta comprobación se usarán los mismos especímenes mostrados en la sección anterior.

Además, se limitó el formato de la cantidad de decimales a solo 3, debido a que los números originales son muy extensos y consideramos que luego de los 3 decimales dicha información es de poca importancia.

Para la primera tabla, se comprobará la invariación de la rotación, donde se harán 4 variaciones, en la 1 se hará con 45°, luego la 2 con 90°, luego la 3 con 135° y por último con 180°.

Para la siguiente tabla, se comprobará la invariación del desplazamiento, donde se harán 4 variaciones, en la 1 se

Vocal	Rotación 1	Rotación 2	Rotación 3	Rotación 4
a	2.800, 21.502	2.800, 21.499	2.799, 21.422	2.799, 21.375
e	2.974, 21.779	2.974, 21.779	2.974, 21.776	2.974, 21.777
i	3.196, -27.323	3.196, -27.398	3.196, -27.366	3.196, -27.364
o	2.848, -24.299	2.848, -24.299	2.847, -24.137	2.847, -24.020
u	2.989, -21.668	2.989, -21.708	2.989, -21.710	2.989, -21.709

desplazará 10 píxeles hacia la derecha, en la 2 se desplazará 10 píxeles hacia arriba, en la 3 se desplazará 10 píxeles hacia la izquierda y en la 4 se desplazará 10 píxeles hacia abajo.

Vocal	Desplazamiento 1	Desplazamiento 2	Desplazamiento 3	Desplazamiento 4
a	2.799, 21.375	2.799, 21.375	2.799, 21.375	2.799, 21.375
e	2.974, 21.777	2.974, 21.777	2.974, 21.777	2.974, 21.777
i	3.196, -27.364	3.196, -27.364	3.196, -27.364	3.196, -27.364
o	2.847, -24.020	2.847, -24.020	2.847, -24.020	2.847, -24.020
u	2.989, -21.709	2.989, -21.709	2.989, -21.709	2.989, -21.709

Para la siguiente tabla, se comprobará la invariación de la escalación, donde se harán 4 variaciones, en la 1 se escalará a 100x100 píxeles, en la 2 se escalará a 150x150 píxeles, en la 3 se escalará a 200x200 píxeles y en la 4 se escalará a 250x250 píxeles.

Vocal	Escalación 1	Escalación 2	Escalación 3	Escalación 4
a	2.799, 21.374	2.799, 21.375	2.799, 21.374	2.799, 21.375
e	2.974, 21.777	2.974, 21.778	2.974, 21.778	2.974, 21.778
i	3.196, -27.357	3.196, -27.356	3.196, -27.355	3.196, -27.355
o	2.847, -24.019	2.847, -24.019	2.847, -24.019	2.847, -24.019
u	2.989, -21.709	2.988, -21.710	2.988, -21.710	2.988, -21.710

### III. RECONOCEDORES

Para todos los reconocedores, cada imagen fue preprocesada usando los métodos de thresholding para eliminar el ruido y facilitar el análisis de la imagen. Luego, se eligieron los bordes blancos con fondo negro. Por lo tanto, la imagen primero se dilata y luego se erosiona. Finalmente, se normaliza el tamaño de todos los especímenes según el más grande y se centran en el cuadro.

#### III-A. Reconocedor de caracteres implementado

Este reconocedor toma el 70 % de los especímenes para estudiar, obtiene el histograma (con la información tanto horizontal como vertical) con el conteo de los píxeles que forma parte de la letra, y luego analiza el 30 % restante de especímenes para identificar la vocal de acuerdo a los valores anteriormente encontrados. Este reconocedor es muy susceptible al escalamiento y la rotación, por eso las imágenes deben tener una buena resolución y someterse al preprocesamiento para ser normalizadas.

#### III-B. Reconocer de caracteres ya existente

El reconocedor utilizado ya existente fue PyTesseract. Según Zelic y Sable (2021), los sistemas de reconocimiento óptico de caracteres se han utilizado ampliamente para proporcionar entradas de texto automatizado en sistemas informáticos. Sin embargo, en todo este tiempo, los sistemas de OCR convencionales (como el OCR zonal) nunca han superado su incapacidad para leer más que un puñado de tipos de letra y formatos de página. [1] Por lo tanto, los reconocedores de caracteres escritos existentes en la comunidad del software

libre funcionan muy bien para identificar letras escritas por una computadora, pero no por una persona, e incluso si son escritas por computadora, son susceptibles al tipo de letra.

#### III-C. Momentos de Hu

Este reconocedor también toma el 70 % de los especímenes para estudiar y utiliza el resto para verificar el correcto funcionamiento. La diferencia entre este y el primero radica en que se obtienen los momentos de Hu en vez del histograma. Los momentos de Hu se utilizan con regularidad para caracterizar la forma de un objeto. Los 7 Momentos Hu son invariantes bajo traslaciones (ya sea que se muevan los ejes x o y), la escala y la rotación. Por lo tanto, los momentos de Hu se puede utilizar para reconocer formas. Si la distancia es pequeña, las formas tienen una apariencia cercana y si la distancia es grande, las formas son más distintas en apariencia. [2], [3] La invariabilidad a la rotación y al escalamiento resulta muy favorable en este estudio para identificar vocales.

### IV. RESULTADOS

Vocal	Porcentaje de aciertos		
	Reconocedor propio	Reconocedor ya existente	Momentos de Hu
a	42.7083	0.0	56.2500
e	40.9091	0.0	60.2273
i	60.0000	1.2780	86.5854
o	55.6701	0.0	73.1959
u	31.2500	0.0	64.5833

### V. ANÁLISIS

Al comparar los tres reconocedores, el más efectivo es el de momentos de Hu. Este tiene muchas ventajas, entre las cuales están su baja dimensionalidad (solo 7 valores), la velocidad con la que se calculan los vectores y el hecho de que no es necesario escalar las imágenes para compararlas. Aunque las imágenes utilizadas ya estaban escaladas y preprocesadas, es muy posible que reconozca las vocales sin necesidad de escalarlas. Cabe destacar, que incluso con los momentos de Hu no se obtienen valores muy altos, esto lo atribuimos al hecho que cada quien tiene su manera de escribir, e incluso nosotros identificaríamos algunas vocales como otra. Por ejemplo, la *a* a continuación puede considerarse una *u*.



Figura 6. Ejemplo de vocal a ambigua

Si bien el reconocedor propio dio resultados considerablemente buenos al identificar vocales, fue más lento que el programa implementado con los momentos de Hu. Esto seguramente por el análisis robusto de cada vocal con histogramas. Como ya se mencionó, el programa de los momentos de Hu fue más veloz porque se comparan menos valores.

Por otra parte, el reconocedor ya existente tuvo pésimos resultados para lo que es identificar letras escritas a mano.

Puede identificar muy bien las letras escritas por computadora, pero las escritas a mano no las reconoce bien. Incluso si logra reconocer que es una vocal, usualmente no acierta la vocal que es. De los pocos aciertos que tuvo este reconocedor, fue la i.

El porcentaje no acertado para la i lo podemos atribuir a algunos especímenes muy distintos a los demás, ya sea por estar escritos en cursiva o en diagonal, que son muy distintos a la mayoría como la figura 3 que es vertical y abarca la mayoría del cuadro.



Figura 7. Ejemplo de vocal i ambigua



Figura 8. Ejemplo de vocal i ambigua

En todos los reconocedores la i fue la vocal más fácil de reconocer debido a su gran diferencia en forma con las demás. En segundo lugar está la o. En tercer lugar, para los momentos de Hu, está la u mientras que esta vocal fue la menos acertada en el reconocedor propio.

Finalmente, el reconocedor propio tuvo una tasa de aciertos por encima del 30 %, si se ignora la u, un 40 %. El porcentaje más alto fue la i con 60 % de aciertos. El de Hu tuvo una tasa de aciertos por encima del 56 %, cuyo valor más alto fue 86.5854 % para la i.

## VI. CONCLUSIONES

Utilizar los momentos de Hu resulta mejor en términos de espacio y tiempo, debido a que solo se usa un vector de tamaño 7, lo que agiliza de manera extraordinaria la evaluación de cada imagen, el obtener las medias y varianzas, y el hecho de guardarlas en un archivo externo. Por lo tanto, se concluye que si el factor temporal y espacial es relevante para el uso de un identificador, la mejor elección también sería la del identificador con los momentos de Hu.

Es fundamental que las fotos se tomen en entornos similares para obtener los mejores resultados. De lo contrario, es extremadamente difícil suprimir el ruido y establecer un estándar para cada vocal. Además, cada imagen debe ser preprocesada para obtener los mejores resultados. Deben tomarse mejores fotografías para identificar las vocales con mayor precisión.

Otro punto importante es que incluso con una buena imagen, es difícil para un programa reconocer las vocales de todas las personas, si hasta para una persona lo es, hay *as* que parecen

*us*, entre otras. Siempre habrá un porcentaje de error, pero este se puede reducir estudiando más especímenes y se obtuvieron resultados bastante satisfactorios por encima de un 56 % en general.

## Resultados del lab 5:

La variación de momentos de Hu dentro de una misma categoría, refiriéndonos a rotación, escalación o desplazamiento, es mínima por no decir que nula. Donde existió mayor variación de momentos de Hu fue en la rotación, aunque está no fue de mayor impacto para confundir una letra con otra, mientras que en el desplazamiento y la escalación no hubo variación dentro de su propia categoría.

Luego, en cuanto a variación entre distintas modificaciones, fue de máximo 0.01 con la vocal i, siendo relativamente pequeña, luego, a parte de la i, las demás vocales tuvieron una variación de 0.001 aún menor a la otra.

Por último, se demuestra que los momentos de Hu sí varían, pero su variación es mínima y no llega a confundir la distinción de una vocal con otra, por lo que permiten utilizarse para identificar distintas modificaciones de una vocal, e inclusive sirven para distinguir hasta cierto punto entre distintas vocales.

## REFERENCIAS

- [1] F. Zelic and A. Sable, "A comprehensive guide to ocr with tesseract, opencv and python," *Nanonets*, 2021. [Online]. Available: <https://nanonets.com/blog/ocr-with-tesseract/>
- [2] M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE Transactions on Information Theory*, pp. 179–187, Feb 1962.
- [3] Z. Huang and J. Leng, "Analysis of hu's moment invariants on image scaling and rotation," *2010 2nd International Conference on Computer Engineering and Technology*, vol. 7, May 2010. [Online]. Available: <https://www.researchgate.net/publication/224146066>