DETECCIÓN DE POSES DE LAS MANOS USANDO DESCRIPTORES LBP

CRISTHIAN DAVID TORRES ABOLEDA

TUTOR: MARCELA IREGUI GUERRERO

FACULTAD DE INGENIERÍA

INGENIERÍA EN MULTIMEDIA

UNIVERSIDAD MILITAR NUEVA GRANADA

2015

Detección de poses de las manos usando descriptores LBP

Cristhian Torres Arboleda
Universidad Militar Nueva Granada
Bogotá, Colombia
Email: cristhian.ta93@gmail.com

Marcela Iregui Guerrero
Universidad Militar Nueva Granada
Bogotá, Colombia
Email: marcelairegui@gmail.com

***Abstract*—El reconocimiento de gestos se ha presentado como una alternativa para la implementación de sistemas de interacción eficaces. Particularmente las aplicaciones basadas en visión artificial poseen ventajas en portabilidad frente a otras alternativas. Sin embargo, los algoritmos suelen requerir entrenamiento computacional intensivo, siendo difíciles de implementar en dispositivos móviles. En este artículo se realiza un estudio preliminar para la detección de poses de la mano usando un algoritmo basado en Patrones Binarios locales, más conocido por su sigla en inglés LBP (Local Binary Patterns). Para lo anterior, se presenta un modelo heurístico de división de la mano en regiones ponderadas diferencialmente, que permite la clasificación directa de los gestos usando una medida de similitud. Los pesos y la distribución de regiones se evaluaron de acuerdo a su precisión en la clasificación de cada pose de la mano. Estas pruebas se realizaron en un conjunto de imágenes, capturadas en condiciones controladas, correspondientes a cinco poses diferentes. El algoritmo propuesto con el esquema de regiones ponderadas muestra una buena capacidad de discriminación y presenta una alternativa válida para futuras aplicaciones.**

Keywords*—*Local binary Pattern (LBP), interacción humano-computador (HCI), reconocimiento de gestos de las manos, pose, visión por computador.

I. Introducción

Este proyecto se enmarca dentro de la problemática existente en términos de interacción con dispositivos electrónicos, ya que, si bien es cierto que el teclado y el ratón son ampliamente usados, no representan una buena opción en términos de usabilidad para aplicaciones multimedia, personas en condición de discapacidad o en equipos móviles. Estos últimos, por ejemplo, a pesar de que cada vez son más potentes, es claro que las limitaciones en el tamaño dificultan la interacción.

La preocupación por mejorar la experiencia del usuario, mediante la oferta de sistemas de interacción natural, ha impulsado la creación de dispositivos y la implementación de algoritmos eficientes para lograr la detección precisa de instrucciones del usuario. Entre las múltiples posibilidades para la interacción por medio de gestos, se han propuesto técnicas basadas en diversos dispositivos como guantes, instrumentos hápticos, marcadores ópticos, controladores con sensores y sensores de movimiento como Kinect entre otros. Si bien muchos de los dispositivos son bastante precisos y conducen a desarrollos rápidos de una amplia variedad de aplicaciones, presentan desventajas relacionadas con los altos costos, la portabilidad y configurabilidad [1], [6].

El reconocimiento de gestos por medio de técnicas de visión por computador presenta una oportunidad para mejorar la interacción en portabilidad y ubicuidad, se han utilizado técnicas basadas en imágenes RGB o RGB-Depth. Estos han sido previamente explorados por medio de diversas técnicas, algunas más efectivas que otras. Trigueiros y otros,[4], evalúan siete algoritmos de reconocimiento de gestos estáticos, basados en imágenes obtenidas con una cámara Kinect. Sin embargo, los resultados presentados no son concluyentes, principalmente por la baja resolución y ruido en las imágenes obtenidas. Por otro lado el amplio uso y acceso a las cámaras web convencionales, que hoy por hoy cuentan con excelente resolución y se encuentran en la mayoría de los dispositivos, presentan una oportunidad más económica y simple para sistemas de interacción basados en visión por computador. En [6], los autores presentan una comparación de diversos métodos de detección de gestos basados en análisis de apariencia o modelos de la mano y en técnicas de análisis de imágenes combinada con aprendizaje de máquina. Sin embargo, el problema de complejidad computacional es un problema importante que tienen la mayoría de este tipo de algoritmos [4], [5]. Por esto la importancia de desarrollar técnicas robustas pero a la vez ligeras, que puedan ser implementadas en dispositivos móviles.

El algoritmo LBP (Local Binary Patterns) ha demostrado ser una excelente alternativa para la caracterización de imágenes por medio del análisis de texturas y para el reconocimiento de patrones, gracias a su rendimiento y eficacia. Por ejemplo, en [2] se plantea un método de reconocimiento de rostros basado en patrones locales binarios, calculados en regiones independientes, las cuales son definidas y pesadas heurísticamente para el proceso de clasificación, por medio de la medida de similitud Chi Cuadrado. Los resultados con el uso de LBP representan avances importantes la para detección de rostros.

El uso de LBP como descriptor para reconocimiento de gestos de las manos ha sido utilizado previamente, combinado con algoritmos de aprendizaje de máquina robustos como AdaBoost en [3], [7] y SVM en [4]. Sin embargo, si bien las tasas de precisión son buenas, los algoritmos de aprendizaje aumentan la complejidad computacional en el reconocimiento.

En este trabajo se realiza un estudio del algoritmo LBP para reconocimiento de gestos de las manos, empleando la caracterización por regiones, de forma análoga a como se propone en [2] para reconocimiento de rostros. Lo anterior usando como único clasificador una medida de similitud basada en ChiCuadrado.

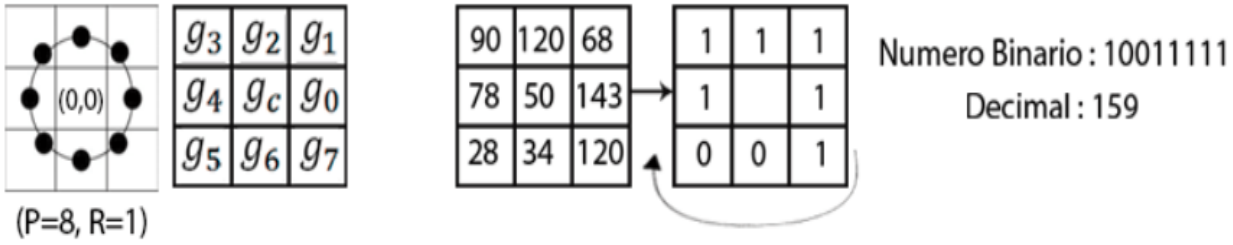

Fig. 1. Operador LBP. A la izquierda, la distribución de la vecindad de 8 píxeles con radio 1. En la parte derecha un ejemplo del cálculo del descriptor para el píxel g_c, en cuyo caso se calculan las diferencias con píxeles vecinos y de acuerdo a la configuración de los cambios, se construye el patrón binario correspondiente

II. Patrones Binarios Locales - LBP

El operador LBP actúa sobre la imagen en escala de grises, a cada pixel se le asigna una etiqueta, de acuerdo con la configuración de la vecindad circular de radio R ver figura 1. Si P es el número total de píxeles vecinos utilizados, g_c es el valor en escala de grises del pixel central y g_p corresponde al valor en escala de grises de un pixel vecino, la etiqueta LBP se calcula según la siguiente ecuación:

$$LBP_{P,R} = \sum_{p=0}^{P-1} 2^p s(g_p - g_c) \tag{1}$$

donde $s(x)$ es la función signo, definida como:

$$s(x) = \begin{cases} 1, x \geq 1 \\ 0, x < 0 \end{cases} \tag{2}$$

Una vez recorridos los pixeles de la imagen o de una región en particular, se calcula el descriptor de textura por medio de la creación de un histograma con 2^P muestras correspondientes a los posibles valores de etiquetas LBP. Dos histogramas de diferentes imágenes pueden ser comparados por medio de una medida de distancia.

III. Método

El método propuesto en este trabajo se basa en el uso de los descriptores LBP y la medida de similitud Chi-Cuadrado, en regiones independientes de la mano. De forma análoga al trabajo presentado en [2], en este artículo se parte de la hipótesis de que, al igual que en las imágenes de rostros, existen zonas características de la mano que aportan información diferencial relacionada con el gesto. Por lo anterior, el método propuesto considera la división de la imagen y la asignación de pesos específicos de acuerdo a una heurística guiada por la concentración de información relacionada con el gesto. Por cada una de las m regiones se calcula un histograma LBP de longitud $n = 2^P$, con lo cual, el número de histogramas calculados es equivalente al número de regiones. Estos se concatenan en un único vector de longitud $m*n$, que representa una descripción global de la imagen.

Teniendo en cuenta la eficiencia que se espera de los algoritmos, no se emplean clasificadores sofisticados. La identificación de los gestos se realiza usando una medida ponderada de similitud chi-cuadrado, de acuerdo con un peso asignado a cada región. Partiendo del principio de que las regiones de la imagen contienen mayor cantidad de información relacionada con el gesto se asigna un peso diferencial a cada región j, ω_j, de tal forma que la comparación entre dos histogramas, H, I , se realiza con la función chi-cuadrado ponderada:

$$\chi_\omega(H, I) = \sum_{i,j} \omega_j \frac{(H_{i,j} - I_{i,j})^2}{H_{i,j} + I_{i,j}} \tag{3}$$

Para determinar la efectividad del clasificador, como primera medida, se analiza el efecto que tiene la comparación independiente de los histogramas de textura, en subdivisiones de la imagen y de acuerdo al tamaño de las mismas. Como segunda medida, por medio de una búsqueda iterativa, se determina la relación óptima de pesos por regiones, con el propósito de identificar la importancia de cada región en la caracterización del gesto.

A. Configuración de las regiones

Con el fin de identificar la capacidad de discriminación del descriptor LBP, en relación con el tamaño de las regiones comparadas, las imágenes se subdividen en partes cada vez más pequeñas y se evalúa el rendimiento de la clasificación. Con el fin de evitar problemas de escala, las imágenes se ajustan al área de delimitación de la mano. (ver figura 2). En este caso a cada región se le da el mismo peso en la ponderación del chi-cuadrado.

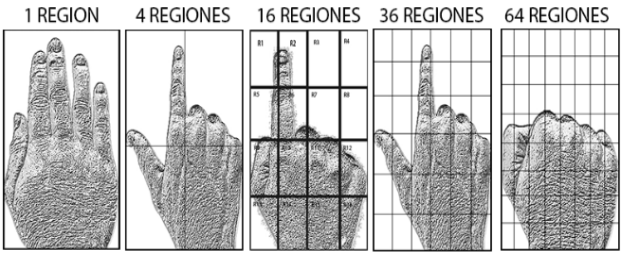

Fig. 2. Configuración de las regiones de acuerdo al tamaño para la comparación del histograma de textura LBP

B. Configuración de los pesos

Una vez analizado el efecto de las subdivisiones, se identifican varias configuraciones de pesos sobre diferentes zonas de la imagen, para evidenciar la importancia de cada una en la identificación del gesto. Lo anterior por medio de un

método heurístico, basado en un protocolo jerárquico que busca establecer las posibles distribuciones de regiones y pesos en diversos niveles de detalle. La figura 3 muestra gráficamente el protocolo, el cual se puede analizar de acuerdo a diversos niveles:

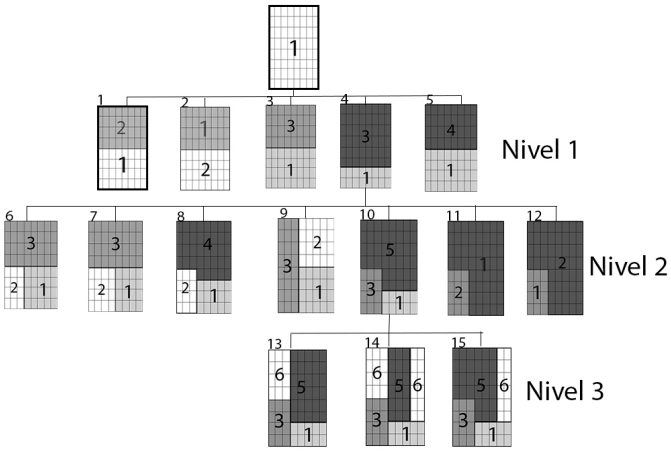

Fig. 3. Niveles de distribución de zonas de la imagen y posibles distribuciones de pesos

- En el primer nivel los pesos se distribuyen en dos regiones. Para este primer caso se separan los dedos (parte superior) y el dorso (parte inferior) y además se busca establecer, que tanto influye la región de los nudillos.
- En el siguiente nivel los pesos se distribuyen en tres regiones, se parte de la configuración con mejores resultados en el nivel anterior y se redistribuyen los pesos con el fin de evidenciar la mejor relación de pesos entre el pulgar, los otros dedos de la mano y el dorso.
- Partiendo de la mejor distribución en el nivel anterior, se subdivide la parte superior de la mano donde se encuentran los dedos para ver en qué grado afecta esto al conjunto del reconocimiento.

IV. Protocolo experimental

Para el desarrollo de las pruebas, se creó una base de datos de 23 imágenes de 5 poses de la mano derecha, para un total de 115 imágenes, puño, palma, índice, pulgar-índice e índice-medio en un ambiente uniforme y controlado, permitiendo la extracción de la mano respecto al fondo y delimitando la imagen a los bordes de la misma. Como paso siguiente, se convierte cada imagen de RGB a escala de grises (ver Fig.4), para ser utilizadas en el análisis del LBP. Todas las imágenes tienen igual tamaño, 400 px de alto y 240 px de ancho.

Para la evaluación se realiza una validación cruzada tipo LOOCV, en la cual se deja una muestra para validación y las restantes como conjunto de entrenamiento. El proceso se realiza para veintitrés (23) iteraciones por cada clase, con cada una de las imágenes de prueba. A partir del conjunto de entrenamiento de cada clase se calcula el histograma $LBP_{P,R}$

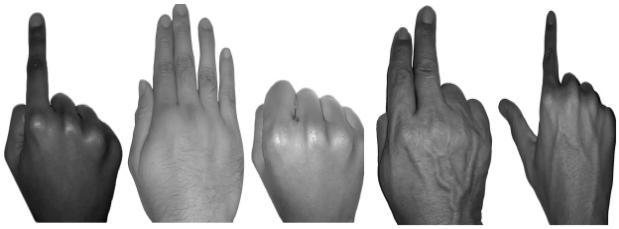

Fig. 4. Las cinco poses que conforman las clases del clasificador

promedio usando como parámetros, una vecindad circular de radio $R = 1$ y un número de ocho píxeles vecinos, $P = 8$.

Para cada imagen de prueba se calcula el descriptor $LBP_{P,R}$, el cual se compara con los histogramas promedio obtenidos en la fase de entrenamiento para cada clase, por medio del chi-cuadrado ponderado. El gesto correspondiente se clasifica en la clase con mayor similitud, es decir, menor valor de χ (Ecuación 3). Lo anterior, se realiza con cada una de las cinco configuraciones presentadas en la sección anterior, para identificar la precisión obtenida en cada caso.

A partir de estos resultados se construye una matriz de confusión que permite identificar el número de Verdaderos Positivos (VP), Verdaderos Negativos (VN), Falsos Positivos (FP), Falsos Negativos(FN), y a partir de estos valores se realizan los cálculos de precisión (PPV), sensibilidad (TPR) y de la medida F (F_1), para cada una de las M clases. Es decir que el cálculo de cada una de estas métricas para una clase i, esta dada por las siguientes ecuaciones.

$$PPV_i = \frac{VP}{VP + FP} \tag{4}$$

$$TPR_i = \frac{VP}{VP + FN} \tag{5}$$

$$F_{1,i} = 2\frac{PPV \cdot TPR}{PPR + TPR} \tag{6}$$

Para la evaluación general de la clasificación, los resultados se reportan según las medidas macro, es decir se promedian cada una de estas tres métricas para el conjunto de todas las clases posibles.

$$PPV = \frac{\sum_{i=1}^{M} PPV_i}{M} \tag{7}$$

$$TPR = \frac{\sum_{i=1}^{M} TPR_i}{M} \tag{8}$$

$$F_1 = \frac{\sum_{i=1}^{M} F_{1,i}}{M} \tag{9}$$

V. Resultados y discusión

Los histogramas promedio de entrenamiento se obtuvieron con veintidós (22) de las veintitrés (23) imágenes de la base de datos en cada pose, la imagen restante es comparada con estos para ser clasificada en una de las cinco poses, por medio del chi-cuadrado ponderado.

De acuerdo con los resultados obtenidos en cada clasificación, se evaluó, para cada distribución en regiones, su precisión, sensibilidad y la medida conocida como medida-F o F-score, para evaluar con un valor único el resultado de las pruebas.

Los resultados de la variación de la precisión y sensibilidad de acuerdo al tamaño de las regiones se muestra en la tabla I. Las cifras muestran que la comparación de los histogramas requiere cierto grado de granularidad, pero que a partir de cierto nivel se comienza a perder precisión. Como se observa se obtiene un valor óptimo de precisión y de F-score al dividir en 64 regiones.

TABLE I. COMPARACIÓN DE LA EFICACIA DEL MÉTODO PARA DIFERENTES SEGMENTACIONES

Regiones	**Distribución WxH**	**Precisión (PPV)**	**Sensibilidad (TPR)**	**Medida-F (F1)**
1 región	1x1	71,40%	69,60%	69,10%
4 regiones	2x2	86,10%	85,22%	85,57%
16 regiones	4x4	87,00%	86,20%	85,80%
36 regiones	6x6	89,50%	88,70%	88,71%
64 regiones	8x8	94,11%	93,91%	93,96%
121 regiones	11x11	93,95%	93,91%	93,91%
256 regiones	16x16	93,95%	93,91%	93,91%
484 regiones	22x22	93,19%	93,04%	93,06%

A continuación se analizan los resultados obtenidos luego de evaluar las diversas configuraciones de pesos. Para esto se parte de una distribución de 64 regiones, por ser esta la que mejores resultados ofrece. Se discute para cada nivel el resultado encontrado de acuerdo con el árbol de la Figura 5.

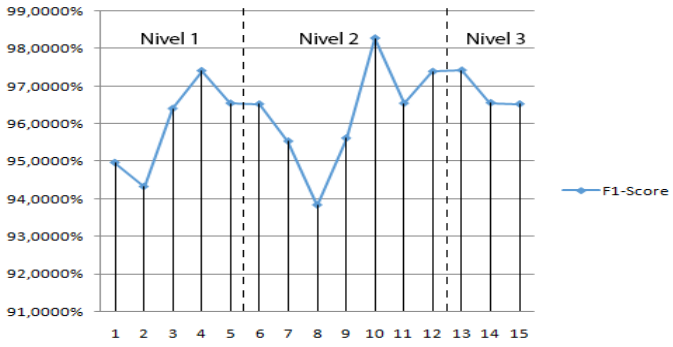

Fig. 5. Medición del F1-score para cada una de las opciones evaluadas, según el árbol de la Figura 3

- **Nivel 1:** El análisis de los resultados presentados en el primer nivel del árbol de la figura 3, demuestran la hipótesis de que los dedos son más discriminantes que el dorso de la mano, en lo que a gestos se refiere. Además, el resultado de la configuración número cuatro permite establecer que los nudillos también contribuyen a la correcta clasificación. Por otro lado se observa que la relación de pesos óptima entre el dorso y los dedos, para este caso es de 1:3 respectivamente.
- **Nivel 2:** Partiendo de los resultados encontrados en el nivel anterior y subdividiendo las imágenes en tres regiones, para identificar independientemente las partes correspondientes al dorso, el pulgar y los cuatro dedos restantes (nudillos incluidos); se encuentran resultados óptimos en la relación 1:3:5 respectivamente (configuración número 10). Evidenciando que las zonas mas relevantes corresponden a los dedos de las manos, siendo el pulgar menos discriminante que los otros.
- **Nivel 3:** De acuerdo con los resultados del nivel 3 se encuentra que realizar mayor cantidad de divisiones para discriminar los dedos de las manos, no representa mayor beneficio en la distribución de los pesos.

Por lo anterior se obtiene una distribución de pesos óptima (configuración 10 de la Figura 3) que permite alcanzar métricas de hasta $98,4\%$ en precisión, $98,26\%$ en sensibilidad y $98,28\%$ en medida-F. Lo cual representa un resultado bastante alentador para el reconocimiento de poses.

VI. CONCLUSIONES Y TRABAJO FUTURO

Se propuso un sistema novedoso para el análisis y reconocimiento de gestos de la mano, basado en la división por regiones de la imagen, lo cual permite analizar cambios de gran relevancia en zonas específicas de la mano. Cada región posee una descripción de la apariencia de una parte de una pose y la combinación de todo el conjunto de regiones dan una descripción global de ésta. Por otro lado, el algoritmo basado en Patrones Locales Binarios (LBP) es un eficiente caracterizador de la pose, robusto y eficaz y dada su baja complejidad computacional, es una excelente alternativa para su implementación en dispositivos móviles.

Finalmente se concluye que de forma similar al rostro, la mano tiene características en regiones que son más determinantes que otras, influyendo significativamente en el reconocimiento de la pose. Es así como en la separación de las zonas superiores y de los nudillos, con valores de peso diferenciales en cada caso, influye en el resultado de similitud existente entre estas mismas.

El número de imágenes de la base de datos, permite aprender mejor de las poses y por consiguiente a su clasificación. Un conjunto más amplio de imágenes de entrenamiento es necesario, permitiendo adquirir más resultados que certifiquen que el LBP es un buen caracterizador de textura y base para la clasificación e identificación de la pose.

Los resultados preliminares se obtuvieron a partir de un conjunto de imágenes de poses, adquiridas en condiciones controladas, para la mano derecha, sin rotación alguna, ni variaciones de iluminación; permitiendo que la división de las regiones de las manos sean similares unas con otras, proporcionando una comparación e identificación más exacta. Si en embargo, es necesario realizar pruebas del LBP para poses mayores factores de variación.

AGRADECIMIENTOS

El presente trabajo es producto derivado del Proyecto de Iniciación Científica PIC ING-1579, enmarcado dentro el Proyecto de Investigación INV ING-1534, financiados por la Vicerrectoría de Investigaciones de la Universidad Militar Nueva Granada - Vigencia 2014

REFERENCES

[1] Meenakshi Panwar and Pawan Singh Mehra, *"Hand Gesture Recognition for Human Computer Interaction"*.International Conference on Image Information Processing (ICIIP 2011). Centre for Development of Advanced Computing Noida, Uttar Pradesh , India.

[2] Timo Ahonen, Abdenour Hadid and Matti Pietikäinen, *"Face Description with Local Binary Patterns: Application to Face Recognition"*. Transactions on pattern analysis and machine intelligence, vol. 28, no. 12, december 2006.

[3] Youdong Ding, Haibo Pang, Xuechun Wu and Jianliang Lan, *"Recognition of hand-gestures using improved local binary pattern"*. International Conference on Multimedia Technology (ICMT), 2011.

[4] Paulo Trigueiros, Fernando Ribeiro and Luís Reis, *"A Comparative Study of Different Image Features for Hand Gesture Machine Learning"*. ICAART 2013. International Conference on Agents and Artificial Intelligence.

[5] Vladimir I. Pavlovic, Rajeev Sharma and Thomas S. Huang, *"Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review"*. IEEE Transactions On Pattern Analysis and Machine Intelligence, vol. 19, no. 7.

[6] Ankit Chaudhary, J. L. Raheja, Karen Das and Sonia Raheja, *"Intelligent Approaches to interact with Machines using Hand Gesture Recognition in Natural way: A Survey"*. International Journal of Computer Science and Engineering Survey (IJCSES) Vol.2, No.1, Feb 2011.

[7] Bin Xiao, Xiang-min Xu and Qian-pei Mai, *"Real-Time Hand Detection and Tracking Using LBP Features"*. School of Electronic and Information Engineering, South China University of Technology Guangdong.