

SISTEMA AUTOMATICO DE RECONOCIMIENTO DE COLONIAS BACTERIAS BASADO EN DIFRACCIÓN LASER

AUTOMATIC RECOGNITION SYSTEM OF BACTERIAL COLONIES BASED ON LASER SCATTERING

Est. Luisa Fernanda Mejía
Institución Universitaria
Salazar y Herrera, Escuela de
Ingeniería, Ingeniería Electrónica
Grupo de investigación GEA.
Cra. 70 #52-49, Medellín,
Colombia,
Lfernanda.14@hotmail.com

Est. Didier Montoya Echavarría
Institución Universitaria
Salazar y Herrera, Escuela de
Ingeniería, Ingeniería Electrónica
Grupo de investigación GEA.
Cra. 70 #52-49, Medellín,
Colombia,
citernontoyag@gmail.com

Miguel A. Becerra
Institución Universitaria
Salazar y Herrera, Escuela de
Ingeniería, Ingeniería Electrónica
Grupo de investigación GEA,
Cra. 70 #52-49, Medellín,
Colombia,
m.becerm@iushedu.co

(Recibido el 09-09-2011. Aprobado el 09-02-2012)

Resumen: Este trabajo presenta un método para la clasificación de colonias de bacterias, basado en naive-bayes para el análisis del espacio de representación generado con medidas estadísticas obtenidas de los momentos de Hue y el método Otsu. Los resultados globales de detección fueron a alrededor del 90%, con validación cruzada 70/30, lo que significa que la capacidad de representación de Hue en combinación con el análisis naive bayes, ofrece un rendimiento aceptable para la clasificación de colonias de bacterias.

Palabras clave: colonia a de bacterias, dispersión laser, Naïve-Bayes.

Abstract: this paper presents a method for the classification of bacterial colonies, based on naive-bayes for representation space analysis generated with statistical measures obtained from the moments of Hue and Otsu method. Detecting the overall results were around 90% with cross validation 70/30, which means that the **representation** capability combined with Hue naive Bayes analysis, provides acceptable performance for the classification of bacterial colonies.

Keywords: bacterial colony, scattering laser, Naïve-Bayes.

I. INTRODUCCIÓN

En microbiología se hace importante la generación de metodologías de caracterización de las bacterias, que permitan la identificación de estos agentes para la toma de decisiones a nivel clínico o industrial, con el fin de optimizar tiempo y recurso. En la actualidad la clasificación de microorganismos se basa en la realización de análisis bioquímicos y moleculares, que presentan un alto costo debido a los requerimientos en equipos y personal especializado, tomando en algunos casos más de cuatro días para llegar a la identificación definitiva del agente, como sucede con las pruebas de detección de ADN/ARN (Lupiz, 2009).

En los últimos años, la eficacia del conteo de células o colonias bacterianas ha sido mejorada por varios esquemas de procesamiento de imágenes automatizados. La literatura reporta diferentes estudios como conteo de colonias bacterianas, serotipificación, análisis morfológico y proteómica/genómica que identifican y clasifican colonias de bacterias en determinadas sustancias o elementos, y en los que se utilizan métodos que permiten observar las características morfológicas de dichas colonias, con la inspección visual de diferentes macroscopios en los que se determinan rasgos considerables como la forma, el grosor y el color (Bae et al., 2007). Padmapriya et al, (2007) se basan en un sistema de difracción laser generando

diferentes patrones de difracción a partir de arreglos de colonias de bacterias en crecimiento sobre una superficie de agar semi-sólido, para la identificación de listeria monocytogenes y otros tipos de especies de listeria, aprovechando las diferencias conocidas de sus caracteres fenotípicos (diferentes expresiones del código genético, igual especie e.g. Humanos son de igual genotipo pero diferente fenotipo).

En investigaciones recientes con dispersión laser, se ha encontrado que un láser dirigido a través de una colonia de bacterias, genera una codificación de sus señales características, lo que genera patrones de dispersión hacia adelante únicos (Bac et al., 2010), por lo cual se han realizado numerosas investigaciones para identificar bacterias patógenas e inoñas, a través de esta técnica denominada dispersión elástica de la luz (ELS) (Robinson et al., 2011) o método de Bardot, pero esta última solo revisa el análisis por colonias independientes, mientras ELS realiza de forma automática el análisis de múltiples colonias en una misma caja de Petri.

Este estudio, un enfoque de clasificación basado en la combinación de Naive-Bayes-Bue de imágenes de colonias generadas por medio de difracción laser, es presentado para proporcionar un mecanismo preciso y objetivo para la clasificación de colonias de bacterias.

2. MATERIALES Y MÉTODOS

2.1 Momentos de *Hu*

Los momentos invariantes pretenden extraer características para poder reconocer los objetos aunque no se encuentren siempre en la misma posición, estén girados o su tamaño sea distinto. Para ello, se parte de los momentos ordinarios donde, para una función continua bidimensional $F(x)$ el momento de orden (p, q) está definido en la ecuación:

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy$$

Así, la imagen digital expresada en (1) se convierte en la ecuación:

$$\mu_{pq} = \sum_x \sum_y (x-\bar{x})^p (y-\bar{y})^q f(x, y) f(x, y) \quad (2)$$

donde, la suma de p y q representa el orden del momento calculado. Para conseguir la invariancia se le resta a las coordenadas de cada punto el centro de gravedad, lo que da como resultado los momentos centrales.

$$\mu_{pq} = \sum_x \sum_y (x-\bar{x})^p (y-\bar{y})^q f(x, y) f(x, y) \quad (3)$$

donde,

$$\bar{x} = \frac{m_{10}}{m_{00}} \quad \bar{y} = \frac{m_{01}}{m_{00}} \quad (4)$$

\bar{x} : centro de gravedad o centroide del objeto en el eje de las abscisas.

\bar{y} : centro de gravedad o centroide del objeto en el eje de las ordenadas.

m_{10} : momento ordinario de primer orden, cuando p es igual a 1 y q es igual a 0.

m_{01} : momento ordinario de primer orden, cuando p es igual a 0 y q es igual a 1.

m_{00} : momento de orden cero - área del objeto.

2.2 Redes Bayesianas

Es una técnica utilizada en clasificación, la cual proporciona un análisis cualitativo de los datos mediante la correlación entre sus atributos y proporciona una medida probabilística de la importancia de estos, lo que la diferencia respecto a otros métodos como árboles de decisión y redes neuronales que no dan una medida cuantitativa de la clasificación. Esta técnica permite estimar las probabilidades, a posteriori, de cualquier hipótesis consistente en el conjunto de datos de entrenamiento, para así escoger la hipótesis más probable. Para estimar estas probabilidades se han propuesto numerosos algoritmos, entre los que cabe destacar el algoritmo Naïve Bayes.

Si se denota $P(D)$ como la probabilidad a priori de los datos, se puede estimar la probabilidad de los datos $P(D|h)$ dada una hipótesis h , con el teorema de Bayes:

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)} \quad (5)$$

Para estimar la hipótesis más probable o *MAP* se utiliza (6), ya que $P(D)$ es una constante independiente de h pero si se asume que todas las hipótesis son igualmente probables se aplica la hipótesis de máxima verosimilitud o *ML*, definida en (7).

$$h_{MAP} = \operatorname{argmax}_{h \in H} P(h|D) \quad (6)$$

$$h_{ML} = \operatorname{argmax}_{h \in H} P(D|h) \quad (7)$$

Así, el clasificador Bayesiano permite determinar la clasificación más probable para un conjunto de datos, los cuales se pueden obtener combinando las clasificaciones de todas las hipótesis aplicables de acuerdo a su probabilidad, obteniendo un valor de clasificación v_j así.

$$P(v_j|D) = \sum_{h_i \in H} P(v_j|h_i)P(h_i|D), \quad (8)$$

y la clasificación óptima está dada por:

$$\operatorname{argmax}_{v_j \in V} \left(\sum_{h_i \in H} P(v_j|h_i)P(h_i|D) \right), \quad (9)$$

De acuerdo con (6) se puede estimar la clasificación de un conjunto de datos con base en el valor más probable, dados los valores de sus atributos coma:

$$v_{MAP} = \operatorname{argmax}_{v_j \in V} P(v_j|a_1, a_2, \dots, a_n) \quad (10)$$

y usando el teorema de Bayes de la ecuación (5) se puede estimar la función de clasificación más probable como:

$$\begin{aligned} v_{MAP} &= \operatorname{argmax}_{v_j \in V} \frac{P(a_1, \dots, a_n)P(v_j)}{P(a_1, \dots, a_n)} \\ &= \operatorname{argmax}_{v_j \in V} P(a_1, \dots, a_n)P(v_j) \end{aligned} \quad (11)$$

Donde, $P(v_j)$ se puede estimar con la frecuencia de las clases, pero para $P(a_1, a_2, \dots, a_n|v_j)$ se tienen muy pocos elementos, por lo tanto el clasificador Naive Bayes (CNB), asume que los valores de los atributos son condicionalmente independientes dada el valor de la clase, es decir:

$$P(a_1, a_2, \dots, a_n|v_j) = \prod_i P(a_i|v_j) \quad (12)$$

Así, el clasificador Naive Bayes se define como:

$$v_{NB} = \operatorname{argmax}_{v_j \in V} \left(P(v_j) \prod_i P(a_i|v_j) \right) \quad (13)$$

donde, los valores $P(a_i|v_j)$ se estiman con la frecuencia de los datos observados. Es importante resaltar que no se hace búsqueda de hipótesis, simplemente se cuentan frecuencias de ocurrencia.

Lo que permite al CNB asumir independencia entre los atributos dada la clase y su estructura por lo que solo se tienen que aprender las probabilidades de los valores de los atributos dada la clase (Morales, 2012).



Fig. 1. Metodología Hue-Naive Bayes

De acuerdo con la Fig. I, el sistema de clasificación consta de 4 etapas así:

3.1 Base de datos

En esta investigación se utilizó una base de datos perteneciente al grupo GEA de la Institución Universitaria Salazar y Herrera (IUSH) y al grupo SINERGIA del Instituto Tecnológico Metropolitano (ITM), de la cual no se brindó información taxonómica ni morfológica para evitar el sesgo de la investigación, en la validación del modelo. Solo se informó que pertenecían a dos clases de colonias de bacterias donde 100 pertenecían a la Clase 1 y 100 a la clase 2.

3.2 Procesamiento

Se parte de las imágenes en RGB a las cuales se le realiza la conversión a binaria y a escala de grises, utilizando los métodos de dilatación y erosión, se genera una rejilla de frontera aplicando el método de Otsu y Hu para hallar el umbral óptimo de cada

imagen por clases y hallar su media con el fin de segmentar las imágenes de cada clase con un mismo umbral, logrando obtener la misma cantidad de información en cada imagen binarizada.

3.3 Caracterización

Después de procesada y segmentada la imagen, se calcularon diferentes momentos estadísticos (media, desviación estándar y varianza) de los histogramas obtenidos de cada una de las capas del color RGB.

3.4 Entrenamiento y clasificación

El análisis estadístico del espacio de características para la clasificación de colonias de bacteria, fue llevada a cabo por un clasificador basado en naive-bayes, el cual fue aprobado mediante validación cruzada 70-30 de 30 fold, donde se analizó la capacidad de representación del espacio de características.

4. RESULTADOS

En la tabla 1 se presenta la precisión del sistema de clasificación de colonias basado en redes bayesianas Naive-Bayes, el cual fue puesto a prueba con diferentes conjuntos de características obtenidos a partir de atributos estadísticos y su rendimiento medido aplicando validación cruzada 70-30 a 200 imágenes alcanzando la tasa de acierto más alta 98% para la clase 2 y 82% para la clase 1, utilizando 24 características.

Tabla 1. Precisión del sistema con diferente número de características

	Bacterial clase 1			Bacterial clase 2		
#Características	9	24	27	9	24	27
Tasa de acierto	0.86	0.82	0.8	0.94	0.98	1

La tabla 2 presenta el rendimiento en términos de la precisión del clasificador con los distintos métodos utilizados para la extracción de características, obteniendo el más alto rendimiento por medio del método de Otsu con 91%.

Tabla 2. Rendimiento de técnicas procesado imagen

Técnica	Tasa de acierto	Tasa de Error
Escala de grises, contraste, Otsu	91%	9%
Extracción de planos RGB	90%	10%
Filtrado (IB*IG) Hu, Hu.*IG	90%	10%
Todas las técnicas sin Hu.*IG	90%	10%
Todas las técnicas	90%	10%

5. CONCLUSIONES

Se obtuvo un sistema para la clasificación binaria de colonias de bacterias a partir de imágenes generadas por difracción laser y espacios de representación generados por medio de atributos estadísticos, para desarrollo de un análisis basado en redes bayesianas, el cual presentó un rendimiento favorable en términos de precisión para la detección de colonias de bacterias, independientemente del tipo. Sin embargo, se presenta una diferencia significativa entre tasa de acierto para detección de la clase 1 respecto de la clase 2, lo cual obtuvo la tasa más alta, por lo que requiere optimizar el espacio de representación mediante selección de características para incrementar la discernibilidad interclase.

RECONOCIMIENTO

Este trabajo se enmarca dentro del proyecto "sistema automático de reconocimiento de colonias de bacterias basado en difracción laser" con código 0805, financiado por la Institución Universitaria Salazar y Herrera.

REFERENCIAS

- Rae, E., Banada, P. P., Huff, K., Bhunia, A. K., Robinson, J. P., & Hirleman, E. D. (2007). Biophysical modeling of forward scattering from bacterial colonies using scalar diffraction theory. *Applied Optics*, 46, 3639-3648.
- Rae, E., Rai, N., Aroonanal, A., Robinson, J. P., Bhunia, A. K., & Hirleman, E. D. (2010). Modeling light propagation through bacterial colonies and its correlation with forward scattering patterns. *Journal of biomedical optics*, 15(4), 045001.
- Banada, P. P., Guo, S., Bayraktar, B., Bae, E., Rajwa, B., Robinson, J. P., Hirleman, E. D., et al. (2007). Optical forward-scattering for detection of *Listeria monocytogenes* and other *Listeria* species. *Biosensors & bioelectronics*, 22(8), 1664-1671.
- Morales, E. (2012). Ciencias Computacionales. Obtenido de Instituto Nacional de Astrofísica, Óptica y Electrónica: <http://ccc.inaoep.mx/-emorales/>
- Robinson, J.P., Rajwa B.P., Bae E, et al. (2011) Using Scattering to Identify Bacterial Pathogens *Optics and Photonics News*, Vol. 22, Issue 10, pp. 20-27