

Reconocimiento de patrones para estructuras cromosómicas

*Luis Enrique Ledezma Fuentes**

RESUMEN

Este artículo propone un método de diseño e implementación de un sistema para la clasificación automática de cromosomas humanos mediante técnica de reconocimiento de patrones. Este sistema es una herramienta que permite ser apoyo para la identificación de posibles síndromes genéticos que ocurren cuando en las células del cuerpo hay un número de cromosomas diferente al de los 46 (23 pares) que corresponden. El tener cromosomas de más o una cantidad inferior constituye una de las causas del desarrollo de algún defecto o síndrome genético. En la primera etapa se realiza un procesamiento de imágenes y una segmentación de los cromosomas la cual es una combinación de técnicas de umbralización local y morfología matemática. Se estudiaron e implementan varios descriptores geométricos y morfológicos representados por un vector de características para la representación de los cromosomas. Se proponen unas redes neuronales artificiales tipo perceptrón multicapa y distancia euclidiana como clasificadores en la etapa de aprendizaje. Los vectores de características y clasificadores se evalúan para dos tareas distintas; reconocimiento de cromosoma vs objetos e identificación de los pares homólogos de cromosomas. De los resultados obtenidos se observa un desempeño aceptable de esta metodología, la cual puede ser empleada para la identificación de aberraciones cromosómicas.

Palabras clave: Cromosoma, cariotipo, procesamiento de imágenes, reconocimiento de patrones, redes neuronales.

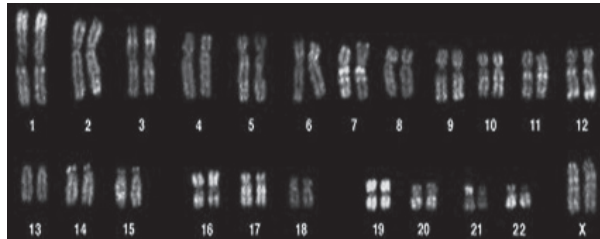
* Facultad de Ingeniería, UAEM. Instituto Nacional de Investigaciones Nucleares. luis.ledezma@inin.gob.mx; luis_enrique1966@hotmail.com

INTRODUCCIÓN

Los cromosomas son organelos constantes en número, forma y características, pero los cromosomas de una célula pueden ser diferentes unos de otros. Ésta diferencia permite clasificar a los cromosomas en metafásicos, metacéntricos, submetacéntricos, acrocéntricos y telocéntricos, según tengan el centrómero en medio, desplazado hacia uno de los brazos, casi en el extremo o en el extremo, respectivamente (Arqueros, 2002).

El cariotipo es un instrumento de vital importancia en el estudio de la citogenética (Córdova, 1997), es un esquema de los cromosomas de una célula ordenados de acuerdo a su morfología y tamaño y es característico a cada especie, al igual que el número de cromosomas que en el caso del ser humano son 46 cromosomas (23 pares porque somos diploides o $2n$) organizados en 22 pares autosómicos y 1 par sexual (hombre XY y mujer XX) tal como se muestra en la figura 1.

Figura 1
CARIOTIPO DE UN SER HUMANO



No obstante, pueden darse casos en los que existen otros patrones a los que se les conoce como aberración cromosómica que es una anomalía de número o estructura de los cromosomas.

El proceso manual de cariotipado de los cromosomas es tedioso, se invierte demasiado tiempo y trabajo. Consiste en revisión visual de los cromosomas en imágenes de metafases (mitosis o división celular) y en obtener medidas o datos de cada una para determinar los 23 pares y elaborar el cariotipo.

Como antecedente se ha desarrollado software comercial para realizar reconocimiento de patrones en cromosomas, por ejemplo, el sistema de cariotipo desarrollado por el Laboratorio de Patología y Centro de

Investigación del Cáncer de la Universidad de Missouri-Columbia cuya limitación esencial es la separación de cromosomas que se encuentran solapadas; también se encuentra el Sistema Cytovision desarrollado por la British Applied Imaging Corporation que no clasifica a todas las cromosomas del cariotipo, lo cual el sistema de cariotipo automatizado desarrollado por el instituto de Visusimaging de Rusia sí realiza. (Wang y Zheng, 2008; Scialdone y Nicodemi, 2007).

En el presente trabajo se muestra el desarrollo de un prototipo de sistema automatizado de identificación de cromosomas en metafase y de cromosomas que presentan algún tipo de rompimiento (aberraciones cromosómicas) mediante técnicas de tratamiento de imágenes y reconocimiento de patrones, el cual está siendo puesto a prueba en el Laboratorio de Biología del Instituto Nacional de Investigaciones Nucleares.

Para llevar a cabo el proceso se han identificado 4 tareas o partes del proyecto:

1. **Adquisición y procesamiento:** Aquí las imágenes son adquiridas haciendo uso de una cámara acoplada a un microscopio óptico. Las imágenes son tratadas mediante uso de filtros espaciales para su mejoramiento y ser utilizadas en etapas posteriores.
2. **Segmentación:** El objetivo de esta fase es segmentar los cromosomas de la imagen de la metafase de entrada. Para ello hay que idear un algoritmo que permita identificar los cromosomas que estén cruzados o solapados y ubicación del centrómero.
3. **Extracción de características:** En esta fase se determinan las características a extraer de las imágenes de los cromosomas segmentados para crear la matriz de características que se empleará como base de conocimiento y entrenar el modelo de clasificación.
4. **Clasificación:** En esta parte del proceso se escoge un método de clasificación que se aplique a la matriz de características y así diferenciar entre cromosoma y una aberración.

Aunque a lo largo del proyecto se analizarán las 4 partes, éste se centrará principalmente en las etapas de segmentación, extracción de características y construcción y entrenamiento de un modelo eficiente de identificación de cromosomas y aberraciones cromosómicas en metafase.

METODOLOGÍA

Un sistema de reconocimiento de patrones completo consiste en un sensor que recoge las observaciones a clasificar, un sistema de extracción de características que transforma la información observada en valores numéricos o simbólicos, y un sistema de clasificación o descripción que, basado en las características extraídas, clasifica la medición (Palacios, 2008).

La metodología utilizada para el reconocimiento de patrones es la descrita por Delie Ming y Jinwen Tian en su artículo “Automatic Patern Extraction and Classification for Chromosome Images” (Ming, 2010); la cual incluye las siguientes fases:

- *Adquisición y procesamiento de imágenes de cromosomas:* Capturar y ajustar a un tamaño de imagen común y definir el tipo de resolución adecuado para observar detalles que sirvan de base para la posterior ubicación de centrómeros en los cromosomas. El procesamiento de imagen consta de aplicación de filtros espaciales (mallas de vecindades de píxel) sobre las imágenes adquiridas con la finalidad de eliminar presencia de ruido que puede presentarse a consecuencia del proceso de captura, digitalización y transmisión propios del sensor de la cámara acoplada al microscopio y mejoramiento de las características de la banda (bandas transversales que permiten definir a cada cromosoma y estudiar su estructura).
- *Segmentación:* Es común observar en las imágenes cromosomas pegadas y superpuestas, o bien, solapadas, que aplicando técnicas de segmentación permiten obtener cromosomas individuales, y así poder realizar extracción de su eje medio o también denominado esqueleto de cromosoma, así como permitir calcular la característica de la banda y evaluar demás rasgos característicos (área, longitud de la cromosoma, por ejemplo).
- *Uso del clasificador:* La metodología sugiere el empleo de un clasificador multicapa mediante una red neuronal con los valores característicos obtenidos para su identificación, adicionalmente se propone el empleo del clasificador de distancia euclidiana para la agrupación de pares de cromosomas adyacentes.

Igualmente se apoyó en la metodología de procesamiento de imágenes digitales propuesta por Rafael C. González y Richard C. Woods (González, 2001): Adquisición, Procesamiento, Segmentación, Descripción y Clasificación y que además aporta el uso de una base de conocimiento que para el presente trabajo se conforma por los valores característicos de las cromosomas.

DESARROLLO DEL SOFTWARE

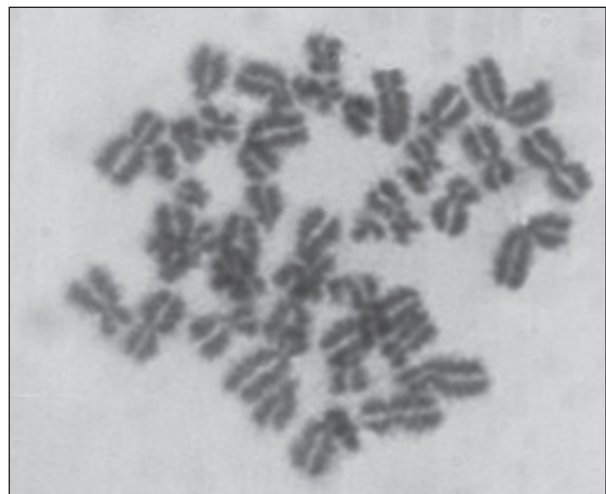
El sistema de reconocimiento de patrones fue desarrollado en Matlab® (abreviatura de MATrix LABoratory, Laboratorio de Matrices) versión 2009a. Es un software matemático que ofrece un entorno de desarrollo integrado (IDE) con un lenguaje de programación de alto nivel propio (denominado lenguaje M). Este software está disponible para plataformas Unix, Solaris, Windows y MacOS.

A. Adquisición de las imágenes

Las imágenes de las metafases celulares se adquirieron en un microscopio Nikon Eclipse E400 a una magnificación de 100X en formato JPG con una resolución de 1280x1024 bajo el estándar de color RGB (Figura 2).

Figura 2

IMAGEN DE CROMOSOMA

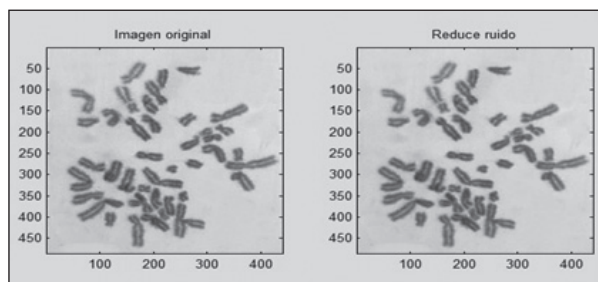


B. Procesamiento

Ming y Tian destacan la importancia de realizar un procesamiento de la imagen previo a la segmentación para realzar la imagen y eliminar el posible ruido presente y principalmente para obtener una mayor cantidad de detalle de los cromosomas. Para este fin se propone la utilización de wavelets para descomponer la imagen, eliminando los componentes grandes, oscuros y redondos y aquellos con un área menor al 1/220 del área total. Una vez eliminado la mayoría del ruido se aplicó un filtro morfológico, por ejemplo el Filtro de Área. Esta clase de filtros morfológicos presenta la ventaja de filtrar los objetos deseados sin alterar la forma original, lo que produce menor distorsión en la imagen (ver figura 3).

Figura 3

IMAGEN ORIGINAL Y POSTERIOR
A LA APLICACIÓN DE FILTRO DE REDUCCIÓN DE RUIDO



34

C. Segmentación

Los métodos más simples para segmentación de cromosomas están basados en umbrales, pero éstos no son suficientes para separar los grupos de cromosomas concentrados. La separación de cromosomas que se tocan sin cruzarse se realiza encontrando una línea de corte entre las dos, para las solapadas en cambio, es necesario encontrar dos pares de corte más o menos perpendiculares (Goinetxea, 2011).

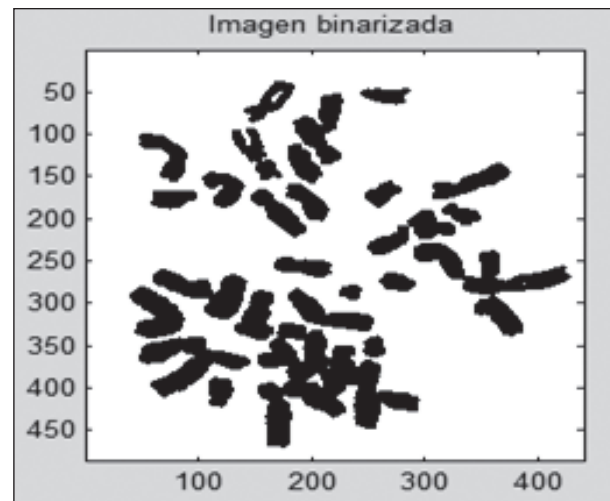
Se evaluaron distintos métodos de segmentación, entre los cuales se destacan método Otsu y Canny para detección de bordes y umbralización local adaptativa, que permitieron obtener los mejores resultados. Para eliminación de posibles hoyos o agujeros en cromosomas se aplica un closing (acercamiento),

consistente en dilatación y posterior erosión, lo anterior redundará en una mejor ubicación del centrómero de la cromosoma y por ende, en un mejor desempeño de la contabilización de pares de cromosomas, medidas de los brazos y detección de aberraciones.

La figura 4 ejemplifica imagen resultante al aplicar los métodos anteriores.

Figura 4

IMAGEN DE CLUSTER DE CROMOSOMAS
DURANTE PROCESO DE SEGMENTACIÓN



Con la imagen esqueletizada se pueden identificar los cromosomas solapados o cruzados analizando la vecindad de cada pixel: Si un pixel tiene dos vecinos con valor 1, ese pixel no hace frontera con otro cromosoma, entonces no se solapan en ese punto. Si el pixel tiene más de dos vecinos con valor 1, ese pixel hace frontera con otro cromosoma, entonces no hay solapamiento (Goinetxea, 2011).

El centrómero es localizado basado en características geométricas y distribución de niveles de gris. L_0, L_1, \dots, L_n es el eje medio de una cromosoma, donde L_i, L_{i+1} ($0 \leq i < n$) es una serie de líneas cortas y $(n - 1)$ es el número de esas líneas cortas. Entonces la longitud de los cromosomas puede ser calculado usando:

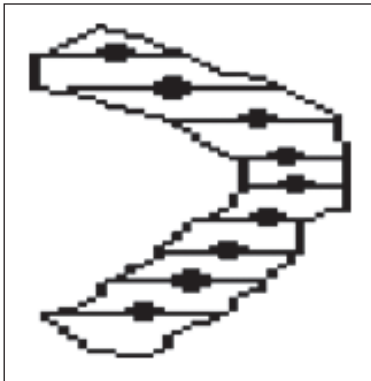
$$|L_{i+1} - L_i|$$

Entonces el radio del centrómero puede ser expresado como:

$$C = \min(L_1, L_2) / L$$

El eje medio es un parámetro para describir la forma y estructura de una cromosoma, el cual es extraído mediante el algoritmo de puntos medios, aquí la imagen es escaneada línea por línea hasta la posición del centrómero en ambos lados y finalmente los puntos medios de cada línea se conectan (Figura 5).

Figura 5
EXTRACCIÓN DEL EJE MEDIO



D. Extracción de características

La extracción del eje medio tiene la importancia debido a que facilita el calcular las características de banda: gradiente y forma del cromosoma, así como promedio de nivel de gris de la cromosoma.

Se evaluaron distintos descriptores geométricos y morfológicos entre los cuales se encontraron área, perímetro y momentos de Hu, además del índice del centrómero (radio entre el brazo corto del cromosoma y su longitud total).

E. Clasificación

Se hace uso del clasificador multicapa o multietapa. En la primera capa los cromosomas son clasificados en diversos grupos considerando sus características geométricas (longitud relativa, área, perímetro, posición de centrómero y característica de banda) posteriormente se implementó el cálculo de la distancia mínima euclidiana debido a que los descriptores como tal no permitían diferenciar un par de cromosomas con

otro, además que en cada metafase cambia la forma y el tamaño de los cromosomas dificultando crear un estándar de tamaños y formas para cada grupo de par.

La figura 6 muestra la agrupación de pares de cromosomas de un individuo sano obtenido al finalizar el proceso de reconocimiento de patrones.

Figura 6

EJEMPLO DE CARIOTIPO DE PARES
DE CROMOSOMAS OBTENIDO A PARTIR DE IMAGEN DE PRUEBA



EVALUACIÓN DEL SISTEMA DE RECONOCIMIENTO DE ESTRUCTURAS CROMOSÓMICAS

Para la evaluación del Sistema de Reconocimiento de Patrones de Estructuras Cromosómicas actualmente se lleva a cabo una comparación entre los resultados emitidos manualmente y los emitidos por el sistema

En imágenes que mostraron irregularidades debido a toma de la muestra u oquedades se obtuvieron los siguientes datos:

Numero de objetos:	57 cromosomas
Numero de objetos por clúster:	2
Numero de grupos:	29 pares

Los resultados anteriores pueden indicar la presencia de aberraciones cromosómicas, sin embargo no fueron avaladas por el personal del Laboratorio de Biología del ININ debiéndose entonces de continuar mejorando el software propuesto.

Para el clasificador multicapa se está probando con 100 imágenes obteniendo un 85.6% de exactitud, mientras que para el clasificador euclidiano con el mismo número de imágenes un 72.3%.

CONCLUSIONES

La eficiencia del sistema se está demostrando, aún faltan pruebas por hacer con imágenes de diferentes muestras de cromosomas ya que las realizadas hasta ahora provienen de cromosomas de un solo individuo sano. No se ha probado con personas que tienen alguna enfermedad derivada de las aberraciones cromosómicas. Se analizaron imágenes con problemas de contrastes con las que se obtuvo una gran tolerancia y buen funcionamiento para cada cromosoma evaluado.

El proceso de segmentación es la base del sistema, su resultado influye en las demás etapas. Por tal motivo se realiza una investigación exhaustiva a diferentes técnicas hasta lograr un híbrido entre la umbralización adaptativa y el método Otsu que optimice la separación de los cromosomas en un 90% o superior.

En el análisis de los descriptores y determinación del vector característico se concluyó que los descriptores de forma (momentos invariantes de Hu) son los que presentan mayor grado de eficiencia en el clasificador ya que los cromosomas son muy similares en su morfología.

PERSPECTIVAS

El sistema de reconocimiento de patrones queda con posibilidades de mejora para su funcionalidad. Algunas recomendaciones para futuros desarrollos son las siguientes:

1. Pruebas con muestras de cromosomas de seres humanos de diferentes edades y sexo, sanos y enfermos.
2. Definir el mejor careotipo que se hubiese obtenido al día de hoy.

3. Involucrar la implementación de una platina móvil con una interface que permita la manipulación desde una computadora para la adquisición automática de imágenes en tiempo real.

REFERENCIAS

- Arqueros, E., *Ponencia: Nuevas Salidas Profesionales: Genética- Citogenética-Genética Molecular*, XIII Congreso Nacional Farmacéutico Granada, 15-18 de octubre de 2002.
- Córdova, J., Lamas, G., (1997). *Citogenética, Filogenia, Clasificaciones Naturales Y Evolución De Las Especies*. Editorial Alma Mater. pp 95 – 112.
- Ming, D., Tian, J., (2010). *Automatic Pattern Extraction and Classification for Chromosome Image*. Journal of Infrared, Millimeter, and Terahertz Waves, 31(7), 866-877. Estados Unidos. Springer.
- Goinetxca, U., (2011) *Diseño y Validación de un Método de Análisis de Imagen para Cariotipado Automático*. Tesis de Maestría. Departamento de Ciencias de la Computación e Inteligencia Artificial. Universidad del País Vasco, España.
- González, R., Woods, R., (2001). *Digital Image Processing*. (2da. Ed.), Madrid, España: Ed. Prentice Hall.
- Palacios, M., (2008). *Reconocimiento y consulta de imágenes textuales en bibliotecas digitales*. Tesis de Licenciatura [Internet], Puebla, Universidad de las Américas. Disponible desde <http://catarina.udlap.mx/u_dl_a/tales/documentos/lis/dirccio_p_r/portada.html>[Acceso 8 febrero 2008].
- Scialdone, A., and Nicodemi, M., (2007). *Pairing of homologous chromosomes as phase transition*. Proceedings of SPIE 6802, 68020F.
- Wang, S., and Zheng, B., (2008). *Development and assessment of an integrated computer-aided detection scheme for digital microscopic images of metaphase chromosomes*. Journal of Electronic Imaging 17, 043008.