



**UNIVERSIDAD CENTRAL "MARTA ABREU" DE LAS VILLAS**  
**VERITATE SOLA NOBIS IMPONETUR VIRILISTOGA. 1948**

**Facultad de Ingeniería Eléctrica.**

**Centro de Estudios de Electrónica y Tecnologías de la  
Información**

## **TRABAJO DE DIPLOMA**

**Base de datos anotada de imágenes de leucocitos  
para la evaluación de algoritmos de procesamiento  
digital de imágenes**

**Autor: Reynier Figueredo Cervantes**

**Tutor: Msc. Alexander Falcón Ruiz**

**Santa Clara**

**2011**

**"Año 53 de la Revolución"**





**UNIVERSIDAD CENTRAL "MARTA ABREU" DE LAS VILLAS**  
**VERITATE SOLA NOBIS IMPONETUR VIRILISTOGA. 1948**

## **Facultad de Ingeniería Eléctrica.**

**Centro de Estudios de Electrónica y Tecnologías de la  
Información**

# **TRABAJO DE DIPLOMA**

**Base de datos anotada de imágenes de leucocitos  
para la evaluación de algoritmos de procesamiento  
digital de imágenes**

**Autor: Reynier Figueredo Cervantes**

**Tutor: Msc. Alexander Falcón Ruiz**

Profesor Auxiliar. Msc., CEETI, Facultad de Ingeniería Eléctrica,

[afalcon@uclv.edu.cu](mailto:afalcon@uclv.edu.cu)

**Santa Clara**

**2011**

**"Año 53 de la Revolución"**

**CON SU ENTRAÑABLE TRANSPARENCIA**





Hago constar que el presente trabajo de diploma fue realizado en la Universidad Central “Marta Abreu” de Las Villas como parte de la culminación de estudios de la especialidad de Ingeniería Biomédica, autorizando a que el mismo sea utilizado por la Institución, para los fines que estime conveniente, tanto de forma parcial como total y que además no podrá ser presentado en eventos, ni publicados sin autorización de la Universidad.

Firma del Autor

Los abajo firmantes certificamos que el presente trabajo ha sido realizado según acuerdo de la dirección de nuestro centro y el mismo cumple con los requisitos que debe tener un trabajo de esta envergadura referido a la temática señalada.

Firma del Tutor

Firma del Jefe de Departamento  
donde se defiende el trabajo

Firma del Responsable de  
Información Científico-Técnica

## *Pensamiento*

*“La mayoría de las ideas fundamentales de la ciencia son esencialmente sencillas y, por regla general pueden ser expresadas en un lenguaje comprensible para todos”*

*Albert Einstein*

## *Dedicatoria*

*A mis padres y mi hermana por ser lo más importante que tengo en la vida.*

*A mi novia Danay Hernández López, por ser mi refugio, mi apoyo y la luz de mi vida.*

*A todas esas personas que confiaron en mí y me dieron fuerzas para seguir adelante.*

## *Agradecimientos*

*A mis padres, por el apoyo que me han dado a lo largo de todos los años de estudio de mi vida.*

*A mi hermana por creer y confiar en mí.*

*A mi novia Danay, por ayudarme a afrontar estos momentos tan difíciles, darme aliento cuando más lo necesitaba, por brindarme su amor y comprensión.*

*A la familia de mi novia, que siempre estuvieron al tanto de mis necesidades y de brindarme aliento para seguir adelante, por ayudarme y apoyarme como si fuera un hijo.*

*A mi tutor Alexander Falcón Ruíz por apoyarme, guiarme y darme confianza para llegar hasta el final.*

*A mis profesores durante estos 5 años de estudios en la Facultad de Eléctrica por todo lo que con paciencia y amor me han enseñado, en especial a los profesores del CEETI.*

*A todas aquellas personas que de una forma u otra contribuyeron a la realización de este sueño hecho realidad.*

## TAREA TÉCNICA

Tareas a realizar para cumplir los objetivos:

1. Estudio del manejo del microscopio trinocular Accu-Scope 3016PL con cámara digital 318CU y del software Micrometrics SE Premium para la adquisición de las imágenes.
2. Estudio del software ImageJ para la anotación de la imágenes adquiridas.
3. Adquirir imágenes de leucocitos y realizar anotación de las mismas.
4. Poner a punto algoritmos de segmentación y comprarlos haciendo uso de la base de datos anotada.



Firma del Autor



Firma del Tutor

## RESUMEN

El análisis de sangre ha adquirido gran relevancia en el diagnóstico clínico ya que constituye una ayuda fundamental en el diagnóstico de enfermedades. En este proceso la digitalización de imágenes y algoritmos automatizados poseen gran ventaja sobre el método tradicional de un especialista observando muestras mediante un microscopio, lo cual consume mucho tiempo y es un trabajo tedioso. Para realizar la puesta a punto de algoritmos de análisis automatizados como los de segmentación y clasificación se necesita de una base de datos de imágenes anotadas sobre la cual se pueda corroborar los resultados. En este trabajo se adquieren imágenes de leucocitos o glóbulos blancos utilizando un microscopio óptico con una cámara digital acoplada. Estas se someten a un pre-procesamiento para eliminar el ruido y los artefactos producidos por la iluminación desigual y suciedad en la óptica. Cada una de estas imágenes mejoradas es anotada por un grupo de observadores que deben segmentar manualmente las células en núcleo y citoplasma. Luego, se propone una estrategia para medir la variabilidad inter-observador y finalmente se propone un método para evaluar algoritmos de segmentación. La base de datos obtenida puede ser de gran ayuda en la evaluación de nuevos algoritmos de segmentación, y con una extensión de la misma puede extenderse los estudios a otros tipos de células.

**Palabras claves:** leucocitos, segmentación de imágenes, bases de datos anotadas, variabilidad inter e intra-observador.



## TABLA DE CONTENIDO

Pensamiento .....	i
Dedicatoria .....	ii
Agradecimientos .....	iii
TAREA TÉCNICA .....	iv
RESUMEN .....	v
INTRODUCCIÓN.....	<b>7</b>
CAPÍTULO 1. REVISIÓN BIBLIOGRÁFICA.....	10
1.2 Microscopio.....	10
1.2.1 Partes del microscopio óptico y sus funciones.....	11
1.2.2 Aplicaciones del microscopio óptico.....	12
1.3 Cámara digital.....	13
1.4 Conexión de cámara digital a un microscopio óptico.....	15
1.5 Preparación de la muestra.....	16
1.6 Adquisición de imágenes.....	18
1.6.1 Métodos para adquisición de imágenes.....	19
1.6.2 Calidad de una microfotografía.....	20
1.6.3 Almacenamiento de imágenes.....	21
1.7 Elementos constituyentes de la sangre.....	22
1.8 Anotación de las imágenes y evaluación de segmentación de imágenes.....	25

---

CAPÍTULO 2.	BASE DE DATOS ANOTADA DE IMÁGENES DE LEUCOCITOS.	27
2.1	Materiales .....	27
2.2	Adquisición de las imágenes. ....	28
2.3	Corrección de las imágenes adquiridas.....	30
2.4	Selección del software para la anotación.....	32
2.4.1	Pasos para la segmentación con ImageJ. ....	33
2.5	Medida inter-observador.....	34
2.5.1	Distancia de Jaccard para medir la variabilidad inter-observador.....	35
2.5.1.1	Distancia de Jaccard por parejas y generalizado.....	35
2.6	Comparación de algoritmos de segmentación. ....	36
CAPÍTULO 3.	RESULTADOS Y DISCUSIÓN. ....	39
3.1	Resultados de la distancia de Jaccard para la medición de la variabilidad inter-observador. ....	39
3.2	Comparación de los algoritmos de segmentación utilizados. ....	41
	Conclusiones .....	43
	Recomendaciones.....	43
	REFERENCIAS BIBLIOGRÁFICAS.....	44
	ANEXOS.....	48

---

## INTRODUCCIÓN

En los últimos tiempos el desarrollo de las microcomputadoras ha permitido un aumento en el uso de las imágenes en el campo médico, dada la diversidad y cantidad de información que las mismas poseen. Esta temática del procesamiento computarizado de imágenes reviste una gran actualidad e importancia social y posee un enorme valor científico-técnico, por los innumerables campos donde el tratamiento de las imágenes es de uso común. Las ventajas de la imagen digital frente a la analógica vienen dadas por las múltiples posibilidades de manipulación que nos ofrece. Además podemos procesar dichas imágenes para obtener información, reconocer, contar, medir tamaño, forma, posición o densidad de determinados objetos. El tratamiento digital de imágenes contempla el procesamiento y el análisis de imágenes. El procesamiento esta referido a la realización de transformaciones y a la restauración y mejoramiento de las imágenes. El análisis consiste en la extracción de propiedades y características de las imágenes, así como la clasificación e identificación y el reconocimiento de patrones.

Las imágenes médicas son utilizadas ampliamente para evaluar la salud humana, realizar diagnósticos y recomendar tratamientos. Determinar el contenido exacto de las imágenes para el cuidado de la salud continúa siendo responsabilidad del personal médico. Debido a su extenso uso, las imágenes médicas son archivadas en grandes cantidades en bases de datos con pocas posibilidades de explotar la información contenida en ellas. El desarrollo de las técnicas para el procesamiento de imágenes y las tecnologías para recuperación de información, permiten procesar las colecciones médicas para almacenarlas automáticamente y permitir acceso a su contenido.

El análisis de sangre ha adquirido gran relevancia en el diagnóstico clínico ya que constituye una ayuda fundamental orientada hacia el diagnóstico de diversas enfermedades como el SIDA, el cáncer o infecciones crónicas. Estas son detectadas actualmente por especialistas que observan y extraen información seleccionando un área de interés en una muestra de sangre periférica o médula espinal, detectando mediante un microscopio

---

diferentes tipos de leucocitos, incrementando la cuenta de cada uno y proporcionando información importante a los doctores en el diagnóstico de tales enfermedades. La digitalización de imágenes y algoritmos automatizados poseen una gran ventaja sobre el método tradicional el cual no deja registro de las observaciones obtenidas, por lo cual, para realizar la puesta a punto de algoritmos se necesita de una base de datos de imágenes anotadas. Estas anotaciones requieren un gran consumo de tiempo y es un proceso tedioso para los observadores que segmentan las imágenes manualmente.

Existen contadas empresas dedicadas a la producción de aparatos que realizan tareas como el tratamiento de muestras sanguíneas, un ejemplo es la empresa privada Siemens que presenta un costo que fluctúa por encima de los \$ 50.000 para un Sistema de Hematología; por estas razones no se tiene información sobre la tecnología que utilizan los distintos software de laboratorio y que obviamente el acceso a éstos tienen un determinado costo. Estos costos tan elevados son uno de los principales motivos de que la mayoría de los laboratorios clínicos de nuestro país no puedan acceder a tales aparatos. Por otra parte el software para análisis de imágenes médicas es una opción para el procesamiento de imágenes como radiologías, ecografías, muestras de sangre, etcétera, que ayuden a agilizar tareas rutinarias como el conteo de células sanguíneas, identificación de anomalías y células cancerosas, entre otros. Se encuentran escasos paquetes de software para el análisis de imágenes, la mayoría de los cuales se centran en procesos industriales y no así de análisis de laboratorios.

Por tales motivos, en el presente trabajo se han adquirido imágenes de leucocitos mediante una cámara digital acoplada a un microscopio trinocular. Estas imágenes poseen un alto nivel de ruido producto a aberraciones de la lente e iluminación desigual, requiriendo rehabilitación mediante algoritmos de pre-procesamiento. Luego, en la creación de la base de datos, las imágenes rehabilitadas han sido anotadas por varios observadores empleando el software ImageJ. Cada observador segmentó todas las imágenes identificando el núcleo y citoplasma de los leucocitos encontrados en ellas. Se ha propuesto medir la variabilidad inter-observador empleando la distancia de Jaccard tanto entre pares de observadores como entre todos los observadores. Por último, a modo de ilustración, se propone utilizar el

---

índice de William para la comparación de los resultados de varios algoritmos de segmentación con las anotaciones realizadas por los observadores.

**Objetivo General:**

Crear una base de datos anotada de imágenes de leucocitos para la evaluación de algoritmos, empleando para la adquisición un microscopio óptico y una cámara digital, así como algoritmos escritos en MatLab para la rehabilitación de las imágenes.

**Objetivo específicos:**

1. Formalizar procedimiento para adquirir las imágenes con la calidad requerida.
2. Determinar algoritmos de pre-procesamiento requeridos para la rehabilitación de las imágenes adquiridas.
3. Determinar el software a emplear y procedimiento a seguir para la anotación de las imágenes adquiridas.
4. Proponer estrategias para medir la variabilidad de las anotaciones y para usar la base de datos.

**Organización del informe.**

Este trabajo está estructurado de la siguiente forma: introducción, capitulario, conclusiones, recomendaciones, referencias bibliográficas y anexos. En el Capítulo 1 se tratan las características morfológicas de los leucocitos, los análisis de laboratorio que se realizan para estudiar los mismos, así como la descripción de las herramientas para la adquisición de las imágenes. El Capítulo 2 describe las herramientas utilizadas para la adquisición, los métodos para el pre-procesamiento de las imágenes y la segmentación manual utilizando el software ImageJ; también se explica el proceso de anotación y uso de la base de datos para la validación de algoritmos de segmentación. El Capítulo 3 muestra los resultados de medir la variabilidad inter-observador y la comparación de varios algoritmos de segmentación automáticos haciendo uso de la base de datos anotada. Por último, en los anexos se pueden encontrar los códigos en Matlab utilizados en la realización de este trabajo.

## **CAPÍTULO 1. REVISIÓN BIBLIOGRÁFICA.**

En este capítulo se describe el funcionamiento y aplicaciones de los instrumentos que realizan la captura de las imágenes. Luego, se describe la adquisición y almacenamiento de las imágenes. Además, se explican los elementos constituyentes de la sangre así como la preparación de la muestra de glóbulos blancos y sus diferentes métodos de conteo. Por último, se explica la anotación y evaluación de segmentación de imágenes.

### **1.1 Microscopio.**

En la medicina, el procesamiento digital de imágenes resulta una herramienta invaluable tanto a la investigación como en aplicaciones clínicas, particularmente en diagnósticos, por lo que contar con la posibilidad de visualizar con la máxima claridad imágenes, permite al médico alcanzar mayor fiabilidad así como eficacia al prescribir un tratamiento.

El microscopio es el instrumento que permite observar objetos que son demasiado pequeños para ser vistos a simple vista. El tipo más común y el primero que se inventó es el microscopio óptico. Se trata de un instrumento óptico que contiene una o varias lentes que permiten obtener una imagen aumentada del objeto y que funciona por refracción [1]. También se le conoce como microscopio de luz, microscopio fotónico (que utiliza luz o "fotones") o microscopio de campo claro. Los microscopios de Leeuwenhoek constaban de una única lente pequeña y convexa, montada sobre una plancha, con un mecanismo para sujetar el material que se iba a examinar (la muestra o espécimen). Este uso de una única lente convexa se conoce como microscopio simple, en el que se incluye la lupa, entre otros aparatos ópticos.

Existen varios tipos de microscopios para diversos estudios como suelen ser: el microscopio simple, compuesto, de luz ultravioleta, de fluorescencia, de campo oscuro, de contraste de fase, óptico, entre otros.

### 1.1.1 Partes del microscopio óptico y sus funciones.

En el microscopio distinguimos un sistema óptico destinado a la iluminación y obtención de una imagen muy aumentada del objeto examinado, y un sistema mecánico (o montura), cuya finalidad es la de sustentar convenientemente los elementos ópticos y los preparados que se examinan (Figura 1.1). La parte óptica fundamental del microscopio está formada por dos sistemas centrados de lentes de aumento o convergentes: el objetivo y el ocular, y por el aparato de iluminación, que facilita y mejora la observación microscópica. A continuación se muestran las partes de este microscopio:

- a) Ocular: lente situada cerca del ojo del observador que capta y amplía la imagen formada en los objetivos.
- b) Objetivo: lente situada cerca de la preparación que amplía la imagen de ésta, lo que significa que es muy importante este elemento del microscopio; es un elemento vital que permite ver a través de los oculares.
- c) Condensador: lente que concentra los rayos luminosos sobre la preparación.
- d) Diafragma: regula la cantidad de luz que entra en el condensador.
- e) Foco: dirige los rayos luminosos hacia el condensador.
- f) Tubo: cámara oscura unida al brazo mediante una cremallera.
- g) Revólver: sistema que coge los objetivos, y que rota para utilizar un objetivo u otro.
- h) Tornillos macro y micrométrico: tornillos de enfoque, mueven la platina hacia arriba y hacia abajo. El macrométrico lo hace de forma rápida y el micrométrico de forma lenta. Llevan incorporado un mando de bloqueo que fija la platina a una determinada altura.
- i) Platina: Es una plataforma horizontal con un orificio central, sobre el que se coloca la preparación, que permite el paso de los rayos procedentes de la fuente de iluminación situada por debajo. Dos pinzas sirven para retener el portaobjetos sobre la platina y un

sistema de cremallera guiado por dos tornillos de desplazamiento permite mover la preparación de delante hacia atrás o de izquierda a derecha y viceversa.

Para evitar las aberraciones geométricas se construyen los llamados objetivos planos o planáticos, lo cual suele estar indicado en el propio objetivo con la inscripción PLAN. Los objetivos que están corregidos para las aberraciones cromáticas se denominan acromáticos (corregidos para el rojo y el azul), semiapocromáticos (corregidos para el rojo y el azul y tienen una mayor apertura numérica) y finalmente los apocromáticos (que son de mayor calidad y están corregidos para el rojo, el azul y el verde) [2], [3].

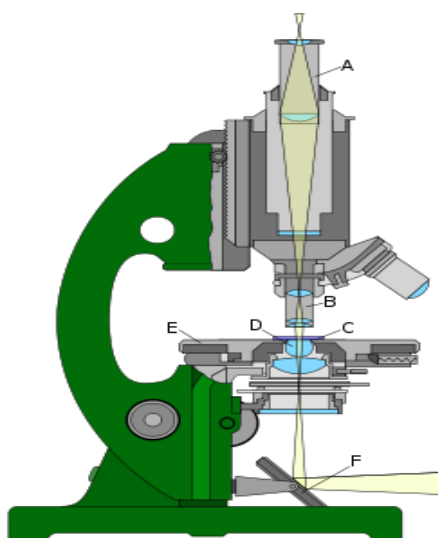


Figura 1.1 Microscopio óptico: A) ocular, B) objetivo, C) portador del objeto, D) lentes de la iluminación, E) sujeción del objeto, F) espejo de la iluminación.

### 1.1.2 Aplicaciones del microscopio óptico.

El microscopio óptico ha sido de gran utilidad en el campo de la medicina, sobre todo en los campos de la ciencia en donde la estructura y la organización microscópica es importante, incorporándose con éxito a investigaciones dentro del área de la química, la física, la geología y en el campo de la biología (en el estudio de estructuras microscópicas de la materia viva), por citar algunas disciplinas de la ciencia.



Hasta ahora se da uso en el laboratorio de histología y anatomía patológica, donde la microscopía permite determinadas aplicaciones diagnósticas, entre ellas el diagnóstico de certeza del cáncer, numerosas estructuras cristalinas, pigmentos, lípidos, proteínas, depósitos óseos, depósitos de amiloide, etcétera [3].

Una variedad de microscopio óptico es el trinocular que a diferencia de los demás posee un tercer ocular para la captura de imágenes mediante una cámara fotográfica acoplada al mismo y que luego son enviadas a una computadora para su análisis.



Figura 1.2 Microscopio óptico trinocular.

## 1.2 Cámara digital.

Una cámara digital es una cámara fotográfica que, captura y almacena fotografías digitalmente mediante un dispositivo electrónico, o en cinta magnética usando un formato analógico como muchas cámaras de video.

La resolución de una cámara fotográfica digital está limitada por el sensor de la cámara (generalmente un CCD o CMOS) que responde a las señales de luz, substituyendo el trabajo de la película en la fotografía tradicional. El sensor se compone de millones de cubos que se cargan en respuesta a la luz. Generalmente, estos cubos responden solamente a una gama limitada de longitudes de onda ligeras, debido a un filtro del color sobre cada uno. Cada uno de estos cubos es un pixel, y se utiliza un algoritmo de mosaicismo e interpolación para unir la imagen de cada gama de longitud de onda por pixel en una imagen del RGB donde están las tres imágenes por pixel para representar un color

completo. Los dispositivos CCD transportan la carga a través del chip hasta un conversor analógico-digital. Éste convierte el valor de cada uno de los píxeles en un valor digital midiendo la carga que le llega. Dependiendo del número de bits del conversor obtendremos una imagen con mayor o menor gama de color. Por ejemplo, si se utilizase un sólo bit tendríamos valores de 0 y 1, y sólo podríamos representar presencia o ausencia de luz, lo que supondría una imagen en blanco y negro puro.

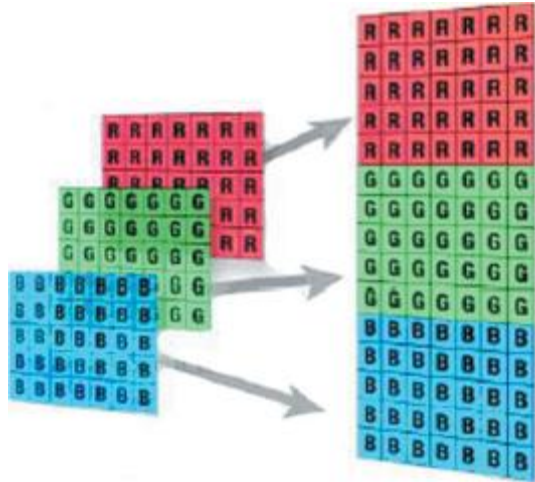


Figura 1.3 Formación de un color.

Por otro lado, los aparatos CMOS contienen varios transistores en cada píxel. El proceso de conversión digital se produce en la propia estructura del sensor, por lo que no se necesita un conversor añadido. Su proceso de fabricación es más sencillo, y hace que las cámaras que utilizan esta tecnología resulten más baratas. La cantidad de píxeles resultante en la imagen determina su tamaño. Existen varios factores que afectan la resolución de un sensor como puede ser el tamaño del sensor, la calidad de la lente y la organización de los píxeles.

Los píxeles excesivos pueden incluso conducir a una disminución de la calidad de la imagen. Cada sensor del píxel es tan pequeño que recoge muy pocos fotones, y así la relación señal-ruido (SNR) disminuirá. Esta disminución conduce a cuadros ruidosos, calidad pobre en sombras y generalmente a imágenes de pobre calidad [4].

### 1.3 Conexión de cámara digital a un microscopio óptico.

Un adaptador óptico mecánico es importante en fotografía digital. Dicho adaptador sirve de enlace entre la cámara y el microscopio. Es importante que la conexión mecánica sea firme, pues cualquier movimiento mínimo o vibraciones de la cámara, reduciría la calidad de la imagen notablemente. Adicionalmente, se requiere un adaptador óptico para el trayecto de luz con el que se logrará así que el sensor CCD/CMOS de la cámara proyecte una imagen de total nitidez e iluminación.

El procedimiento para adaptar la cámara al microscopio consiste en desmontar el lente más exterior del ocular del microscopio y colocar la cámara ensamblada con los accesorios adecuados en lugar de este, sujetándola con tornillos de presión. En la Figura 1.4 se muestra este procedimiento.

Para enfocar la imagen se mueven manualmente las perillas de los micrómetros del microscopio hasta tener una visión clara de la imagen en el monitor. No obstante, existen variaciones en la iluminación que no se logran controlar [5].

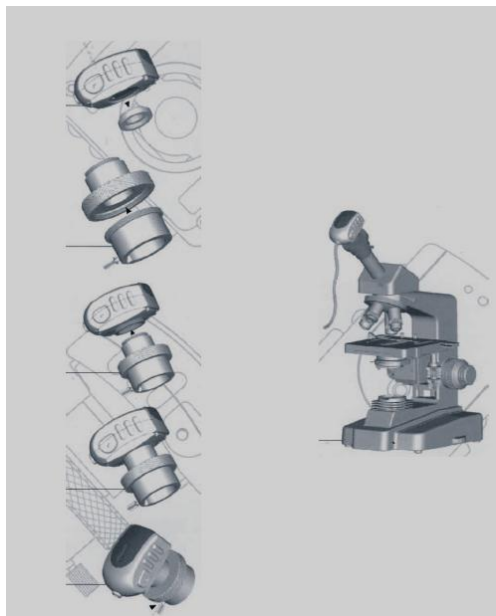


Figura 1.4 Procedimiento de instalación de la cámara. Izquierda.- Montaje de Accesorios. Derecha.- Cámara ajustada al microscopio.

#### 1.4 Preparación de la muestra.

El conteo de los glóbulos blancos es realizado con el fin de identificar los síntomas de alteraciones en el sistema inmunológico que presenta un paciente [6]. Para dicho conteo existen dos métodos [7], diferencial de leucocitos (Figura 1.6) y recuento de leucocitos (Figura 1.7). En el primero, se coloca una gota de sangre fresca sobre un portaobjetos y se extiende con ayuda de otro porta objetos para formar una capa delgada que se deja secar (Figura 1.5), luego se tiñe con azul de metileno glóbulos blancos dándoles una coloración diferente para facilitar su identificación [5],[8]; en el segundo método se le aplica un sustancia llamada líquido de Turk, el cual destruye tanto los glóbulos rojos como las plaquetas de la sangre, dejando sólo los glóbulos blancos. Estos dos métodos nos brindan distintas muestras, las cuales son tomadas por medio del microscopio óptico y analizado para realizar un algoritmo capaz de contar y encontrar el área de los glóbulos blancos.



Figura 1.5 Preparación de frotis sanguíneo.

- Diferencial

Consiste en examinar la morfología de glóbulos blancos, así como su tamaño y número [8]. Este método se caracteriza porque 4.5 ml de la muestra de sangre es mezclada con 0.5 ml de anticoagulante, esta mezcla es agitada manualmente durante tres minutos o en un mezclador. Se agrega azul de metileno para colorar los glóbulos blancos, con el objetivo de su fácil identificación [9],[10].

Este método consiste en realizar un conteo de 100 leucocitos en forma diferencial. Es utilizado para la distinción de los leucocitos.

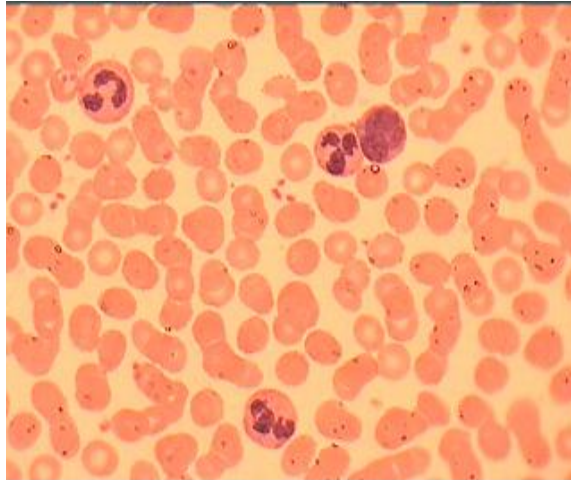


Figura 1.6 Muestra de sangre por el método diferencial de leucocitos.

- Recuento

En este método una muestra de sangre de 0.5ml es mezclada con 0.5ml de Líquido de Turk por medio de pipeta especial, llamada Pipeta de Blancos. El líquido de Turk se utiliza para destruir los glóbulos rojos y las plaquetas de la sangre permitiendo que solo se observen los glóbulos blancos [9],[10]. Para resaltar el color de los leucocitos el líquido de Turk se alteró con Azul de Metileno.

El conteo de los glóbulos blancos se realiza por medio de la cámara de blancos que se encuentra formada por cuatro cuadrículas de 4x4. El número total del conteo de los glóbulos blancos en la cámara es multiplicado por 50, para poder obtener un promedio de los glóbulos blancos por cada mm<sup>3</sup> de sangre. El valor estándar se debe encontrar entre 5.000 y 10.000 leucocitos por mm<sup>3</sup> en un adulto [5] y si la cuenta esta fuera del rango, significa que la persona presenta sintomatología [7].

Normalmente cuando se aplica este método, los leucocitos no son tinturados y dificulta su identificación. Incluso es necesario tener algo de experiencia para realizar el conteo manual. Por esta razón el líquido de Turk es alterado para que los leucocitos presenten una tonalidad que permita al software identificarlos fácilmente [11].

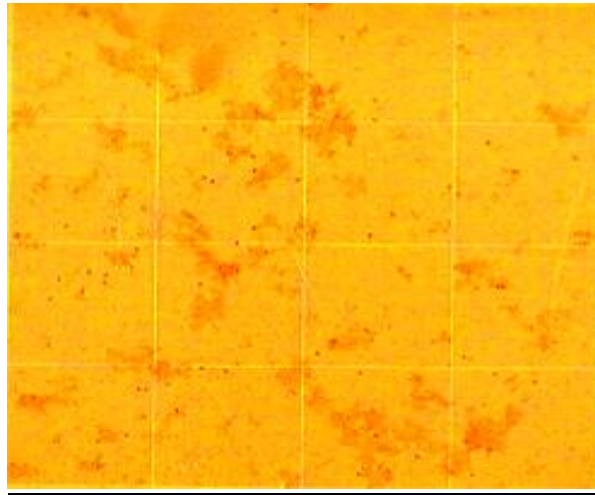


Figura 1.7 Muestra de sangre por el método recuento de leucocitos.

### 1.5 Adquisición de imágenes.

Las placas obtenidas mediante los métodos descritos anteriormente son llevadas al microscopio óptico. Por medio del puerto USB este es conectado al computador para adquirir las imágenes que permiten al programa realizar el conteo e identificación de leucocitos dependiendo del método utilizado (Figura 1.8). Para darle claridad a la imagen, se utiliza un aceite de inmersión sobre la placa.

El microscopio óptico permite enfocar la imagen en distintas resoluciones. Dependiendo del método utilizado se requiere un lente de distinto aumento. Las placas tomadas por medio del método diferencial, son observadas con claridad con un lente de 40x o 100x y las muestras obtenidas por el método recuento de leucocitos son observadas con un lente de 10x (Figura 1.9); estas resoluciones permiten un fácil reconocimiento de los glóbulos blancos. Para implementar el método diferencial se utilizan estas lentes ya que se busca diferenciar 100 núcleos entre monomorfonucleados y polimorfonucleados, no se utilizan lentes de menor aumento. Para implementar el método de recuento de leucocitos la morfología del núcleo es poco relevante, lo que permite utilizar un lente que los identifique pero que no los caracterice.

Por tal motivo la implementación de este método se lleva a cabo con un lente de 10x, ya que este permite tomar imágenes a una cuadrícula de 4x4 perteneciente a la cámara de

recuento de blancos. Obteniendo una tolerancia en el conteo comparable con instrumentos de conteo de leucocitos que se encuentran en el mercado [9],[10].



Figura 1.8 Microscopio óptico con placa de recuento de blancos.

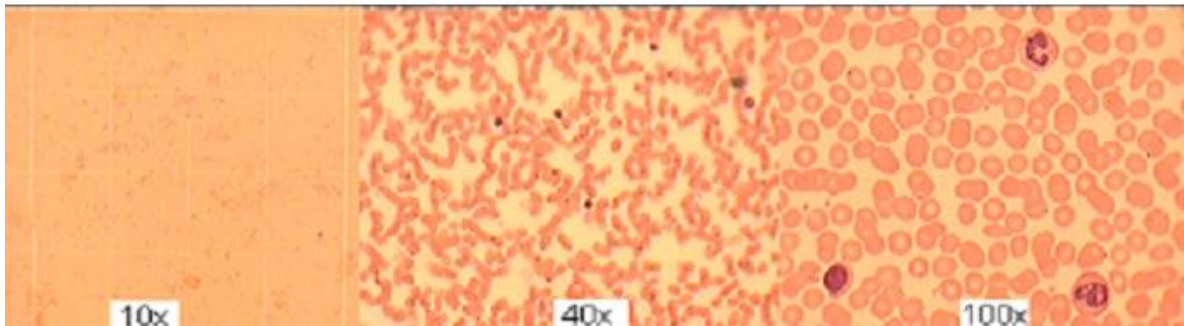


Figura 1.9 Calidad de imágenes en distintas resoluciones [11].

### 1.5.1 Métodos para adquisición de imágenes.

En aplicaciones clínicas, al analizar muestras en un microscopio óptico la manipulación manual para ajustes en el foco se convierte en una tarea intensa y tediosa. El campo de vista en un microscopio se reduce a medida que incrementa el aumento utilizado, de manera que el usuario se ve obligado a mover repetidamente la platina para poder enfocar una nueva porción de la muestra [12]. En la adquisición de las imágenes existen 3 métodos fundamentales:

- El primer método se denomina de disparo único, en referencia al número de veces que el sensor de la cámara fotográfica se expone a la luz que pasa a través de la lente
- El segundo método se denomina de multidisparo, porque el sensor se expone a la imagen en una secuencia de tres o más aperturas del obturador de la lente.

- El tercer método se llama exploración porque el sensor se mueve a través del plano focal como el sensor de un explorador (*scanner*) de escritorio. Sus sensores lineares o tri-lineares utilizan solamente una sola línea de fotosensores, o tres líneas para los tres colores [13].

La elección del método para una captura dada, es determinada en gran parte por el tema a ser fotografiado. Es generalmente inadecuado intentar fotografiar un tema que se mueva con cualquier cosa que no sea un sistema de disparo único. Sin embargo, con sistemas de exploración o multidisparo, se obtiene la más alta fidelidad de color y tamaños y resoluciones más grandes. El soporte lógico inalterable en la cámara fotográfica, o un software en un programa raw del convertidor tal como cámara fotográfica del adobe raw, interpreta las informaciones en bruto del sensor para obtener una imagen completa del color, porque el modelo del color del RGB requiere tres valores de la intensidad para cada pixel: uno por cada uno para el rojo, el verde, y el azul (otros modelos del color, cuando están utilizados, también requieren tres o más valores por el pixel). El patrón del filtro de Bayer es un patrón de repetición del mosaico 2×2 de filtros ligeros, con verde unos en las esquinas opuestas y rojo y el azul en las otras dos posiciones. La parte elevada de verde se aprovecha de características del sistema visual humano, que determina brillo [14].

### 1.5.2 Calidad de una microfotografía.

Las herramientas de procesamiento digital de imágenes que pueden ser utilizadas en las imágenes capturadas con un microscopio óptico, entre ellas están:

Frecuencia de muestreo: Hace corresponder la resolución óptica y electrónica de un microscopio y el sistema de cámara que lo acompaña, una imagen digital debe tener un número suficiente de muestras por la fila horizontal de esta manera el *display* representa fielmente la señal original.

Resolución espacial: Se refiere al número de pixel utilizado en la construcción de la imagen digital. Considerando que las imágenes con un alta resolución son creadas con un mayor número de pixeles que las de baja resolución.



Sustracción del fondo: La aplicación del adecuado algoritmo de sustracción de fondo es una técnica muy útil para corregir los defectos de la imagen asociadas con el brillo no uniforme, a menudo (pero no siempre) atribuido a una iluminación poco uniforme.

Balance Blanco y Negro: Proceso para ajustar la ganancia de color a fin de que los colores de la muestra sean reproducidas correctamente. Este balance ejecuta una estadística de los valores rojo, verde y azul de todos los píxeles en el más reciente marco de la cámara, e invierte la distribución puesta de la ganancia de color apropiadamente. Note que áreas demasiado negras y brillosas en la imagen viva se excluye del cálculo con una exactitud asegurada.

Se deben tener en cuenta tres puntos para realizar este balance:

- Basar el balance blanco en áreas de brillo moderado.
- Las operaciones de balance blanco no pueden cambiar la ganancia del color verde. Si se desea cambia el brillo global de la imagen mientras ejecuta el balance blanco, se tiene que colocar la ganancia del verde automáticamente.
- La ganancia de color debe estar dentro de un rango permitido. Si la ganancia de color es más bajo que el límite permitido la operación del balance blanco es muy probable que sea fallida. Para remediar esto se debe modificar la ganancia de verde consecuentemente y repetir la operación [11].

### **1.5.3 Almacenamiento de imágenes.**

Las fotos se almacenan en ficheros en varias extensiones como JPEG, BMP o bien en formato TIFF o RAW para tener una mayor calidad de imagen pese al gran aumento de tamaño en el archivo. Casi todas las cámaras digitales utilizan técnicas de compresión para aprovechar al máximo el espacio de almacenamiento. Las técnicas de compresión suelen aprovecharse de dos características comunes en las fotografías.

Muchas cámaras fotográficas, especialmente las cámaras profesionales, utilizan el formato raw (crudo). Una imagen raw está formada por el conjunto de píxeles sin procesar obtenidos directamente del sensor de la cámara fotográfica. Editar imágenes en formato

raw permite una mayor flexibilidad en ajustes tales como modificar el balance de blancos, compensar la exposición y cambiar la temperatura de color [14].

### 1.6 Elementos constituyentes de la sangre.

En esta tesis se trabaja con imágenes de microscopía celular, específicamente con imágenes de leucocitos; estos pertenecen a la sangre que es una sustancia líquida que circula por las arterias y las venas de los organismos, la cual está compuesta por tres tipos básicos de células o elementos celulares: los eritrocitos también conocidos como glóbulos rojos o hematíes, las plaquetas y los leucocitos también nombrados glóbulos blancos. La sangre es una forma especializada del tejido conjuntivo, compuesta por una sustancia intercelular líquida llamada plasma.

Los glóbulos blancos presentan distinto aspecto morfológico (Figura 1.10) dependiendo en qué etapa de su vida se encuentren, razón por la que lleva una denominación distinta al encontrarse en una de sus cinco etapas de formación hasta convertirse en un leucocito maduro [16].

Los glóbulos rojos (eritrocitos o hematíes) presentan un color amarillo verdoso pero en masas densas adquieren un color rojo, debido a la alta concentración que contienen de hemoglobina. Este pigmento se separa con facilidad de los hematíes por un fenómeno conocido con el nombre de hemólisis. Presentan la forma de discos bicóncavos y de perfil se presentan como cuerpos alargados con extremos redondeados.

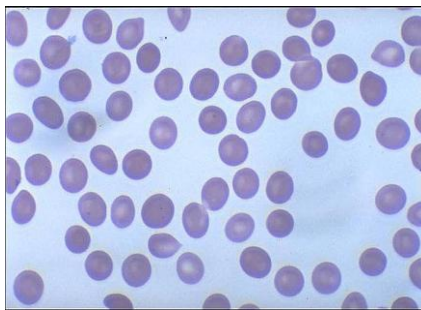


Figura 1.10 Glóbulos rojos [17].

Etimológicamente los leucocitos deben su nombre a las palabras leuco (blanco) y cito (célula). En personas saludables, el número de leucocitos varía entre 5000 y 10 000, en

condiciones patológicas pueden estar aumentados (leucocitosis) o disminuidos (leucopenia) [18].

Los glóbulos blancos o leucocitos son células nucleadas (Figura 1.11) que se encuentran en cantidad mucho menor que los eritrocitos. En la sangre humana pueden distinguirse dos tipos principalmente: los leucocitos agranulosos y los granulosos. Hay dos tipos de leucocitos agranulosos, los linfocitos, que son células pequeñas de tamaño aproximado al eritrocito, núcleo redondeado y escaso citoplasma, y los monocitos, células de mayor tamaño.

Existen tres clases de leucocitos granulosos, los cuales contienen gránulos específicos en su citoplasma. Se les denomina neutrófilos, eosinófilos y basófilos, según la reacción de coloración de sus gránulos citoplasmáticos.



Figura 1.11 Glóbulo blanco.

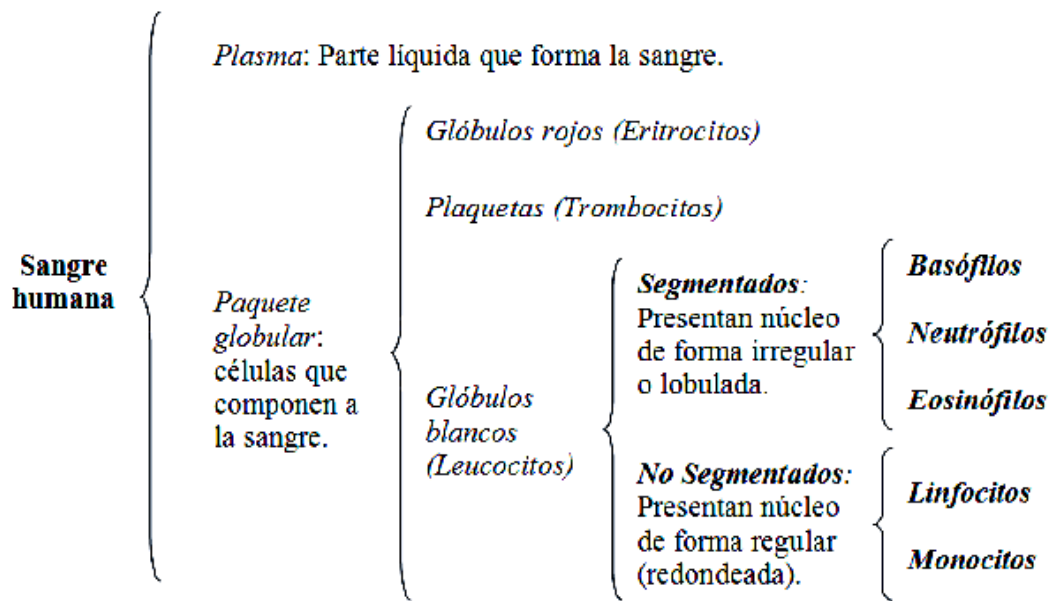


Figura 1.12 Componentes de la sangre humana [5].

Los linfocitos son células esféricas que alcanzan un diámetro de 6-8  $\mu\text{m}$ . Forman parte del 26-40 % de los leucocitos y se presentan generalmente como células redondeadas, de núcleo grande, rodeado por un escaso borde citoplasmático.

Los monocitos están agrupados dentro de los leucocitos agranulosos. Son células de gran tamaño que miden de 9-12  $\mu\text{m}$  de diámetro, comprenden solamente del 2-8 % de los leucocitos de la sangre normal. Poseen un citoplasma abundante de color azul grisáceo pálido.

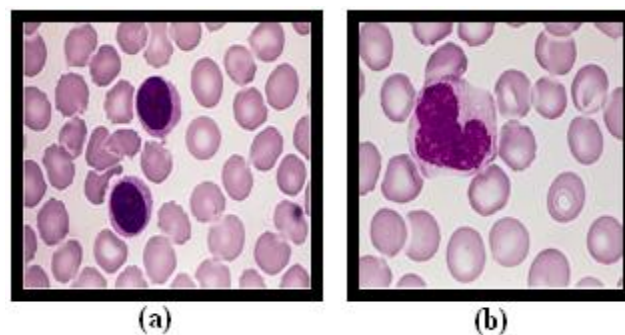


Figura 1.13 (a) Linfocitos, (b) Monocitos [17].

Leucocitos granulados: comprenden del 55-65% del total de los leucocitos y su diámetro varía de 10-15  $\mu\text{m}$  en estado fresco, este tipo de célula recibe su nombre según los numerosos gránulos neutrófilos que abundan en su citoplasma.

En estado fresco tienen aproximadamente de 9-10  $\mu\text{m}$  de diámetro, mientras que en los *frotis* secos varían de 12-14  $\mu\text{m}$ . Representan del 1-3% del total de leucocitos en sangre normal.

Son las células más difíciles de observar, pues constituyen el 0-1% y su tamaño es aproximadamente igual al de los neutrófilos, de 10- 12 $\mu\text{m}$ . El núcleo es de contornos irregulares y en ocasiones bilobular [19].

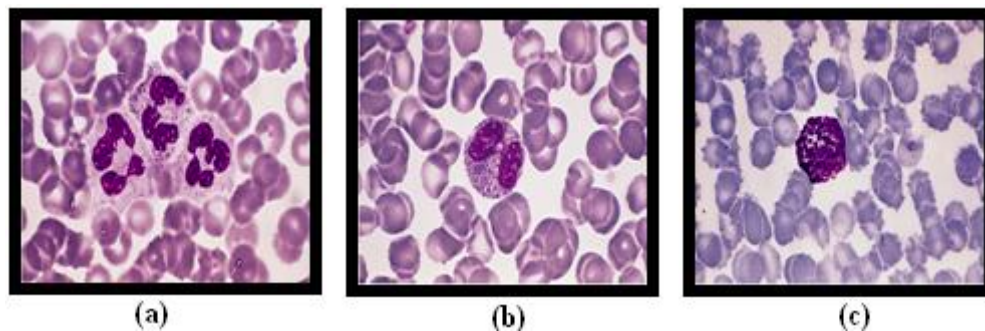


Figura 1.14 (a) Neutrófilos, (b) Eosinófilos, (c) Basófilos [17].

### 1.7 Anotación de las imágenes y evaluación de segmentación de imágenes.

El tipo de interacción requerido por los métodos de segmentación puede ir desde la delineación manual completa de una estructura anatómica, hasta la selección de uno o más puntos iniciales para un algoritmo. La diferencia entre estos tipos de interacción es la cantidad de tiempo y esfuerzo requerido, así como la cantidad de entrenamiento requerido por el usuario. De cualquier forma los métodos de segmentación automatizada generalmente requieren alguna interacción para especificar parámetros iniciales que pueden afectar el rendimiento significativamente.

Este procedimiento posee varios pasos de dibujo o anotación, por lo que es de gran consumo tiempo y llega a ser un proceso tedioso. Los pasos se muestran a continuación:

El trazado manual de un solo volumen de imágenes, puede tardar alrededor de 15 minutos a dos horas o más, dependiendo del número de imágenes en el volumen y el número de anotaciones por la imagen. El perfilado manualmente, es muy dependiente de la opinión del observador, por lo que se producen resultados que están sujetos a prejuicios y errores. Los estudios han mostrado coeficientes de variación de 5% o más para los trazados repetidos de

las mismas estructuras por un solo observador, y de 50% en algunos casos para los trazados de estructuras por múltiple observadores. La gran variabilidad de los resultados de tal análisis casi es inútil, para una necesidad de mejorar la aclaración de los puntos, con un método de medida preferentemente automatizado. También requiere de algún conocimiento del operador con respecto a lo que desea segmentar en la imagen. Estas limitaciones reducen la ventaja de precisión proporcionada por el trazado manual, introduciendo la opinión subjetiva en el proceso de identificación. Además reduce la ventaja de velocidad requerida del extenso trabajo del operador [20].

La primera decisión para la evaluación de segmentación de imágenes es elegir un parámetro a comparar como el área cerrada o el perímetro [21],[22] de la segmentación realizada. Una vez elegido el parámetro a comparar, los observadores segmentan todas las imágenes manualmente y luego se segmenta de manera automática las imágenes adquiridas mediante el microscopio con cámara digital acoplada. La vía más común para evaluar el resultado de la anotación de un grupo de expertos [23] (anotado o segmentados a mano) consiste en compararlos con los resultados de la segmentación efectuada automáticamente. Esta comparación de los resultados de la segmentación de contorno de un sólo observador con la establecida por la computadora no es suficiente, porque el límite de un solo observador puede estar sujeto al prejuicio y a la variabilidad del inter-observador. Existen metodologías para comparar los límites generados por la computadora con los de múltiples observadores, para verificar si los límites generados por esta, difieren de los límites perfilados manualmente, como muchos límites perfilados a mano pueden diferir uno de otro [24].

## CAPÍTULO 2. BASE DE DATOS ANOTADA DE IMÁGENES DE LEUCOCITOS.

En este capítulo se describe el proceso de adquisición de imágenes de leucocitos mediante un microscopio óptico con cámara acoplada. Luego, se explica la fase de pre-procesamiento de las imágenes con el objetivo de mejorar la calidad de las mismas, eliminando errores producto a la iluminación desigual, aberraciones de la lente del microscopio, entre otros. Después, se describe el uso del software ImageJ para la segmentación manual de los leucocitos en núcleo y citoplasma, obteniendo una base de datos anotada. Por último, se muestra cómo puede ser utilizada la base de datos para la validación de algoritmos de segmentación.

### 2.1 Materiales

Para el desarrollo de este trabajo se utilizaron varios materiales en el proceso de adquisición de las imágenes. Las mismas fueron obtenidas con una cámara digital modelo 318CU la cual posee un sensor micrométrico CMOS con el que se pueden obtener imágenes de colores de 24 bits [25]; la cámara posee las propiedades siguientes:

Tabla 2.1 Propiedades de la cámara digital 318CU.

Propiedades	Características
Área de imagen (mm x mm)	6.55 x 4.92
Tamaño del pixel ( $\mu\text{m}$ x $\mu\text{m}$ )	3.2 x 3.2
Tipo de formato óptico	$1/2$
Relación S/N (dB)	43
Imagen de salida	3x8-bit
Resolución (H X V)	2048 x 1536
Software compatible	Micrometrics SE Premium

La cámara fue acoplada a un microscopio trinocular Accu-Scope 3016PL (Figura 2.1) con 10X de oculares, el mismo está compuesto por 3 objetivos que poseen diferentes grados de aumento 10x, 40x y 100x, estos números corresponden a la escala en la que se observan los objetos a través del ocular. Para distinguir adecuadamente las imágenes digitalizadas se utilizó el objetivo de 100x, colocando una gota de aceite de inmersión entre el porta objetos que contiene el frotis sanguíneo y la lente del objetivo. Para la adquisición de tales imágenes se empleó el software Micrometrics SE Premium, compatible con la cámara digital.



Figura 2.1 Microscopio trinocular Accu-Scope 3016PL con cámara digital 318CU acoplada.

## 2.2 Adquisición de las imágenes.

Micrometrics SE Premium es un software para capturar una imagen de alta calidad (bajo nivel de ruido), es eficiente mediante la explotación de la función de imagen fija de cámaras Micrometrics <sup>TM</sup>. Cada cámara Micrometrics <sup>TM</sup> es capaz de integrar fotones entrantes por un tiempo muy largo (cientos de segundos para los modelos X90), además de ofrecer imágenes en directo a una velocidad constante.

En adquisición de las imágenes con calidad se tuvo en cuenta el brillo general, la reproducción del color y el nivel de ruido. Teniendo en cuenta los factores de ajuste único, el brillo general de una imagen ocurre por el nivel de iluminación, tiempo de exposición y ganancia.

Para los efectos de eficiencia, la iluminación debe ajustarse en primer lugar. Mientras que un brillo general deseable no puede lograr solo el ajuste de la iluminación, luego se ajustó



la exposición, a partir del requisito constante máximo permitido por la velocidad de fotogramas. El ajuste de la ganancia debe ser intentado, como último recurso, porque un valor de ganancia mayor que la unidad, amplifica el ruido así como el brillo de la imagen.

Micrometrics SE Premium proporciona la herramienta de automatización Balance de blancos, esta realiza una estadística de los valores de rojo, verde y azul de todos los píxeles en la mayoría de los últimos marcos de la cámara, e invierte la distribución mediante el establecimiento de las ganancias de color adecuadamente. Se tiene en cuenta que las zonas demasiado oscuras y brillantes en la imagen se excluyen del cálculo para asegurar la exactitud.

El balance de blancos es el proceso para ajustar el color de las ganancias para que los colores de la muestra se reproduzcan correctamente. Se puede configurar los colores directamente en las ganancias, pero no es fácil llegar a una juste correcto de esta manera. Existen 3 aspectos que se deben tener en cuenta en las operaciones de balance de blancos:

- Base de la operación de balance de blancos en las áreas de brillo moderado. Los píxeles oscuros pueden introducir ruido y los píxeles saturados pueden no reflejar la información de color verdadero en todos.
- Las operaciones automatizadas de balance de blancos no pueden cambiar la ganancia de color verde. Si se desea cambiar el brillo general de la imagen mientras se realiza el balance de blancos, se debe de ajustar la ganancia manualmente.
- La ganancia de color debe estar dentro del rango permitido. Si una ganancia de color está en su parte inferior o superior, limita la operación de balance de blancos y probablemente falle. Para remediar este fallo se debe modificar el aumento de verde en consecuencia y repetir la operación de balance de blancos.

Las imágenes estáticas (exposición larga) es el modo que se destina a la captura de imágenes de alta calidad (poco ruido o nada introducidos por las grandes ganancias) mientras que el modo de imagen de vídeo (alta velocidad de cuadro) es de pre-visualización de la imagen.

Si la imagen capturada es para efectos de visualización y medición y no sujetos a otras de filtrado en la linealidad de los valores de píxeles, puede ser empleada gamma para lograr el

brillo deseado de la imagen. Un valor gamma que es menos que la unidad aumenta rasgos oscuros y suprime las características claras de la imagen. Un valor gamma que es mayor que la unidad suprime rasgos oscuros y brillantes mejora características de una imagen. Gamma es una transformación no lineal de los valores de pixel, originalmente introducido en Micrometrics SE Premium para compensar la gamma de una analogía tradicional de cámara de vídeo y la de un dispositivo de visualización (microscopio).

### 2.3 Corrección de las imágenes adquiridas.

En las mejores condiciones de imagen, la iluminación a través de un campo celular no es del todo uniforme. Esto se debe a las imperfecciones (es decir, ligero desajustes, el polvo y las propiedades físicas adicionales) de cada elemento óptico (lámpara, filtros, espejos, objetivos) dentro de una luz en la ruta de acceso.

En general, estos errores son particularmente evidentes en las imágenes digitales con fondos brillantes y uniformes, que son producidas por una variedad de modos comunes de iluminación del microscopio incluyendo, contraste de fase oblicua, y el contraste de interferencia diferencial (DIC).

La sustracción de fondo (*background subtraction*) es una técnica que da lugar a alteraciones localizadas de cada valor del pixel en la imagen en bruto, dependiendo de la intensidad de un pixel correspondiente en la misma coordenada de ubicación en la imagen de fondo. Como resultado se obtiene, faltas de uniformidad en la sensibilidad del detector o iluminación que puede ser compensado mediante el almacenamiento de una imagen de fondo de un campo microscópico vacío como una imagen de referencia.

La corrección de campo plano es un enfoque de uso común para este problema. Para esta corrección (también llamada corrección de sombreado), la imagen de un campo blanco (que no contiene material celular o fluorescente) se recoge utilizando como si se estuviera recogiendo imágenes de las células pero en un área vacía de un cubre objetos en el mismo plano focal donde se encuentran las células. La imagen negra se obtiene retirando la cámara del microscopio y se coloca un paño negro en el lente. Es entonces la corrección de campo plano, una división sobre una base pixel por pixel, de cada imagen en bruto por su campo conjugado en blanco con re-escalado [26].

Las imágenes modificadas por la corrección de campo plano (*flat field correction*) parecen similares a las obtenidas con la extracción de fondo, pero realiza la operación de división. La principal razón de esta diferencia es el resultado de la amplitud de la luz de las imágenes que afecta los valores obtenidos por un proceso de multiplicación que combina el flujo luminoso y tiempo de exposición. Después de la aplicación de técnicas de corrección de campo plano (pero no necesariamente los algoritmos de extracción de fondo), las amplitudes relativas de las características de muestra son fotométricamente precisas. Como beneficio adicional, la corrección de campo plano (*FFC*) elimina la mayoría de los defectos ópticos que están presentes en la imagen (Figura 2.2) en bruto mediante la ecuación (1).

$$I_{corregida} = (I_{cruda} - I_{oscura}) / (I_{blanca} - I_{oscura}) \quad (1)$$

La aplicación de técnicas de corrección de campo plano para imágenes digitales a menudo puede garantizar la precisión fotométrica y eliminar los defectos comunes de imagen para restaurar la fidelidad de las características y lograr un equilibrio visual. Con el fin de compensar el ruido y baja intensidad, los marcos de campo plano de referencia pueden estar expuesto por períodos más largos que los utilizados para la captura de imágenes en bruto [27]. Una vez que los marcos necesarios han sido recogidos, la corrección de campo consiste en varias funciones secuenciales. En primer lugar, el marco oscuro maestro se resta tanto de la imagen en bruto y los marcos de campo plano de referencia, seguido por la división de los valores resultantes.

En efecto, el marco de primas se divide por el marco de campo plano después de que el marco oscuro ha sido restado de cada fotograma y el cociente se multiplica por el valor del pixel, significa el fin de mantener la coherencia entre las intensidades de la imagen en bruto y la corregida. Los pixeles individuales en la imagen corregida se ven obligados a tener un valor de nivel de gris entre 0 y 255 [28].

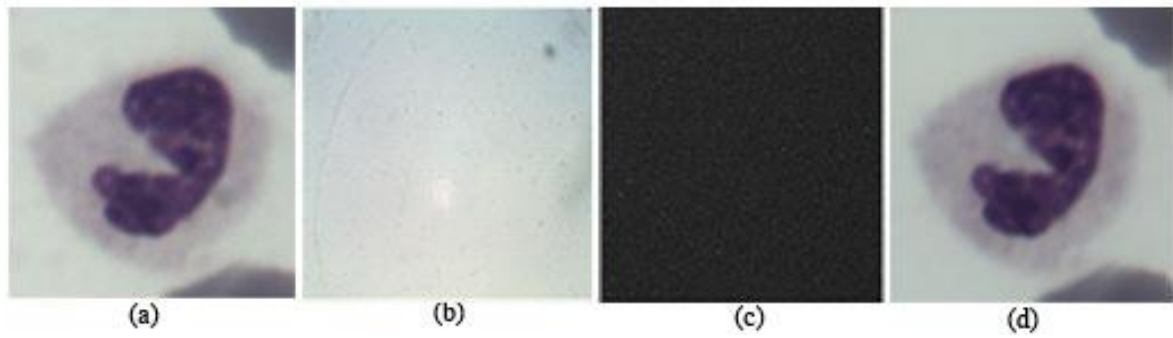


Figura 2.2 Corrección de campo plano (*flat field correction*) de una imagen digital, (a) imagen cruda, (b) imagen de campo claro, (c) imagen oscura, (d) imagen corregida.

## 2.4 Selección del software para la anotación.

Existe una gran variedad de herramientas para la anotación manual de las imágenes, dentro de los que se encuentra el ImageJ [29] el cual se utilizó por su flexible manipulación, su amplio soporte de formato y mayor conocimiento sobre el mismo. El proceso de anotación se efectúa tomando una imagen de leucocito y segmentándola manualmente según la apreciación y conocimiento previo del observador. La segmentación se realiza tanto al núcleo como al citoplasma; el observador no es capaz de lograr la misma anotación en una imagen reiteradas veces por lo que se produce un error de medición; posteriormente estas imágenes anotadas son analizadas por un experto.



ImageJ: es un programa multiplataforma [30], está escrito en Java. Es compatible con imágenes de 8, 16 o 32-bits y soporta los formatos TIFF, GIF, JPEG, BMP, DICOM, FITS y RAW. Además es multitarea y ofrece un panel de funciones muy sofisticado que va desde la colorimetría, análisis de contornos, filtros de imágenes, cálculos de áreas, ángulos y distancias entre píxeles, histogramas de densidad, y mucho más ya que también soporta operaciones estándares de procesamiento de imágenes como por ejemplo la manipulación del contraste, la detección de contornos, el filtrado de medianas, el aumento de nitidez o efecto de suavizado, etc.

Calcula con gran facilidad valores de los píxeles así como áreas específicas, previa detección de contornos. Puede medir distancias y ángulos, pero también ejecutar transformaciones geométricas (rotar, escalar, etc.) y *zoom* para ampliar la imagen y efectuar

un mayor análisis. ImageJ puede trabajar con un número ilimitado de imágenes simultáneamente.

#### 2.4.1 Pasos para la segmentación con ImageJ.

Veinte observadores fueron empleados para el proceso de anotación de un conjunto de datos de 48 imágenes. A los observadores se les instruyeron para dibujar una línea continua entre el núcleo y el citoplasma, y entre el citoplasma y el fondo para cada una de las células (Figura 2.4). Para realizar la anotación de las imágenes en el *software* ImageJ (Figura 2.3) se muestran a continuación los siguientes pasos:

- Cargar la imagen en el programa ImageJ; esta se amplía mediante el *Zoom* para observar adecuadamente los detalles de dicha imagen .
- Seleccionar la herramienta *Polygon* para segmentar primeramente el núcleo y luego el citoplasma marcando manualmente con el *mouse* todo el borde de los mismos .
- Seleccionar en la barra de herramientas *Edit* donde se despliega una serie de opciones de herramientas dentro de las cuales se encuentra *Selection*, en dicha opción se escoge *Fit Spline* para suavizar las esquinas abruptas que marcamos manualmente, al seleccionar esta herramienta aparecen automáticamente puntos sobre la curva trazada que nos sirven para corregir la marcación con el fin de mejorar los resultados.
- En esa misma opción se escoge *Create Mask* con el objetivo de crear una máscara binaria de la imagen original anotada para llevar a cabo la segmentación.
- Una vez obtenida la nueva imagen es necesario invertir la misma, esto se efectúa en el menú *Edit* con la opción *Invert*.

Se pueden obtener en ocasiones células que poseen el núcleo dañado, los cuales son segmentados individualmente y luego con el software Matlab se llevan a binarias y posteriormente se suman para unir los núcleos en una sola imagen.

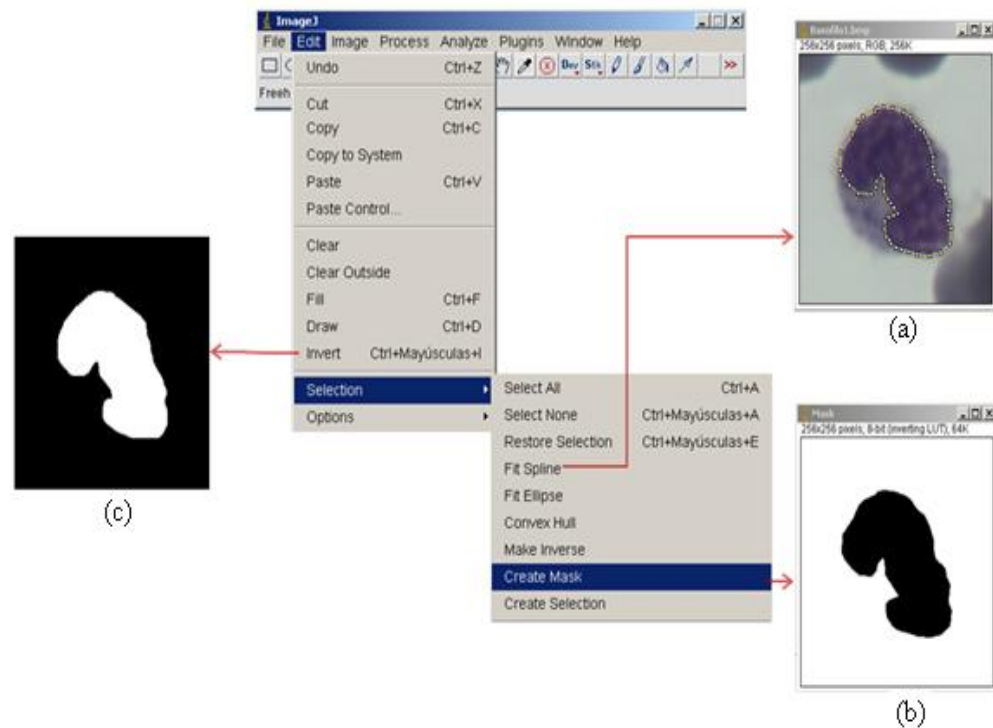


Figura 2.3 Procedimiento para la segmentación manualmente en ImageJ, (a) imagen anotada, (b) máscara, (c) imagen invertida.

## 2.5 Medida de variabilidad inter-observador.

A la hora de realizar la segmentación de las imágenes ocurre que debido al efecto de filtrado paso bajo de la óptica utilizada para capturar las imágenes la región por donde se debe segmentar cubre un número de píxeles, y cada observador puede tomar distintas decisiones (variación inter-observador). También influyen las habilidades de cada observador, su estado de ánimo, los dispositivos de visualización empleados y un gran número de factores más. Incluso un mismo observador puede realizar segmentaciones distintas sobre una misma imagen (variación intra-observador)[31]. Los expertos humanos poseen una excelente capacidad de reconocimiento, debido al conocimiento global que poseen, sin embargo, su capacidad de delimitación es pobre. La variabilidad intra-observador en algunos trabajos se ha reportado alrededor del 6,5%, mientras que la inter-observador en el orden del 22,1%, según [32],[33].

En este trabajo se propone utilizar la distancia de Jaccard para medir la variabilidad inter-observador.

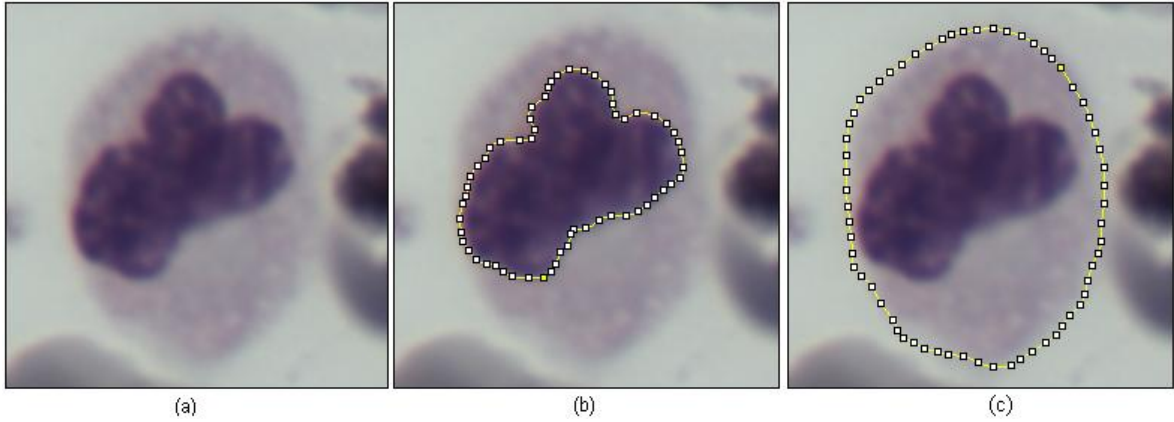


Figura 2.4 (a) Imagen original, (b) anotación del núcleo, (c) anotación del citoplasma.

### 2.5.1 Distancia de Jaccard para medir la variabilidad inter-observador.

El índice o distancia de Jaccard, es un método usado para comparar la similitud y diversidad de muestras y en el caso de imágenes binarias (como las obtenidas en el resultado de la segmentación) podría definirse como la cardinalidad de la intersección dividido por la cardinalidad de la unión de las dos imágenes segmentadas según la siguiente fórmula:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (2)$$

donde A y B son las imágenes sobre las cuales se calcula la distancia y  $|\cdot|$  es la cardinalidad (cantidad de puntos blancos o unos en la imagen binaria).

Ahora bien, en nuestro caso en el que disponemos de varias segmentaciones realizadas sobre una misma imagen (una hecha por cada observador), debemos proponer un modificación. Una primera variante podría ser calcular para una imagen determinada la distancia de Jaccard sobre todos los pares de observadores, ofreciendo el valor promedio como medida de la variabilidad en la segmentación sobre esta imagen en particular. Otra variante podría ser, para una imagen determinada, calcular la cardinalidad de la intersección de todas las segmentaciones dividido la cardinalidad de la unión de las mismas.

## 2.6 Uso de la base de datos anotada para la comparación de algoritmos de segmentación.

Los algoritmos son generalmente evaluados haciendo una comparación de la segmentación realizada por la computadora con la efectuada manualmente. Chalana [24] empleó una modificación del índice de Williams. En el índice de Williams se divide el número medio de acuerdos entre el ordenador y un observador por el número medio de acuerdos entre los observadores [34]. Si el valor superior del intervalo de confianza del resultado es mayor que la unidad, entonces se concluye que la segmentación realizada por el ordenador es confiable a partir de la comparación con el grupo de observadores.

El índice de Williams modificado propuesto por Chalana puede calcularse como

$$I' = \frac{\frac{1}{n} \sum_{j=1}^n \frac{1}{D_{0,j}}}{\frac{2}{n(n-1)} \sum_j \sum_{j': j' \neq j} \frac{1}{D_{j,j'}}} \quad (3)$$

donde en el numerador se calcula la suma ponderada de los inversos de las distancias entre las segmentación realizada por la computadora y los  $n$  observadores, mientras que en el denominador aparece la suma ponderada sobre todas las parejas de observadores, para una misma imagen.

Para estimar el intervalo de confianza (IC) puede usarse la técnica no paramétrica de muestre jackknife [35]. En este procedimiento trabaja dejando fuera una de las  $N$  observaciones cada vez y calculando el índice de Williams para las  $N-1$  observaciones, obteniendo los índices  $I'_{(1)}, I'_{(2)}, \dots, I'_{(N)}$ . El método mencionado estima el error estándar mediante

$$sc = \left\{ \frac{1}{N-1} \sum_{i=1}^N [I'_{(i)} - I'_{(.)}]^2 \right\}^{1/2} \quad (4)$$

donde



$$I'_{(.)} = \frac{1}{N} \sum_{i=1}^N I'_{(i)} \quad (5)$$

Así, el 95% IC para el estimado del índice de Williams modificado es

$$I'_{(.)} \pm Z_{0.95} SC \quad (6)$$

donde  $Z_{0.95} = 1.96$  es el 95% percentil de la distribución estándar normal.

Para el cálculo de distancia se fundamenta una métrica,  $c(\mathcal{A}, \mathcal{B})$ , entre las dos curvas dadas,  $\mathcal{A}$  y  $\mathcal{B}$ . Cuando los puntos de referencia están disponibles en las dos curvas, el establecimiento de la correspondencia entre las curvas es sencillo. Sin embargo, en la ausencia de puntos de referencia, primero tenemos que establecer correspondencias artificiales entre los puntos en las dos curvas y luego medir la distancia entre los correspondientes puntos.

Si las dos curvas se representan como conjuntos de puntos (ecuación 7 y 8), donde cada uno  $a_i$  y  $b_i$  es un par ordenado coordenadas de  $(x, y)$  de un punto en la curva, se define la distancia al punto más cercano (DCP) para  $a_i$  de la curva  $B$  como:

$$A = \{a_1, a_2, \dots, a_m\} \quad (7)$$

$$B = \{b_1, b_2, \dots, b_m\} \quad (8)$$

$$d(a_i, B) = \min_i \|b_j - a_i\| \quad (9)$$

La distancia de Hausdorff entre las dos curvas se define como el máximo de los DCP entre las dos curvas.

$$c(A, B) = \max[\max_i \{d(a_i, B)\}, \max_j \{d(b_j, A)\}] \quad (10)$$

El punto más cercano a la distancia de cada punto en las dos curvas hasta el punto de la otra curva, la distancia de Hausdorff encuentra la mayor distancia entre los puntos asociados. La Figura 3.1 muestra dos ejemplos de la distancia de Hausdorff entre dos curvas, donde es moderada a lo largo de la línea gruesa. En ambos ejemplos es aproximadamente 35 pixeles, considerando que la distancia entre los puntos correspondientes en el primer ejemplo es aproximadamente diez pixeles y en el segundo ejemplo es aproximadamente 30 pixeles. La distancia entre las dos curvas es una métrica si cumple las tres propiedades siguientes:

$$c(A, A) = 0 \text{ and } c(A, B) \geq 0 \quad (11)$$

$$c(A, B) = c(B, A) \quad (12)$$

$$c(A, C) \leq c(A, B) + c(B, C) \quad (13)$$

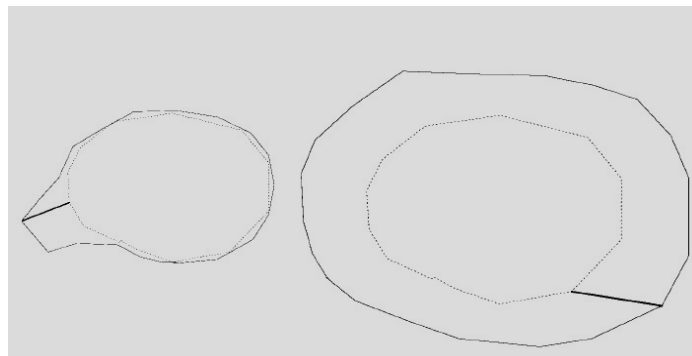


Figura 3.1 Distancia Hausdorff entre las dos las curvas.

## **CAPÍTULO 3. RESULTADOS Y DISCUSIÓN.**

En este capítulo se muestran y comentan los resultados obtenidos al medir la variabilidad inter-observador con la distancia de Jaccard a partir de las segmentaciones realizadas manualmente por los observadores. Adicionalmente, a modo de ilustración de cómo emplear la base de datos anotada, se comparan tres algoritmos de segmentación automatizados de uso extendido (Otsu [36], Kmeans [37] y Contornos Activos [38],[39],[40]) empleando para ello el índice de Williams.

### **3.1 Medición de la variabilidad inter-observador empleando de la distancia de Jaccard.**

Con el objetivo de estimar cuan parecidas entre sí son las anotaciones realizadas por los observadores sobre las distintas imágenes de la base de datos, se ha empleado la distancia de Jaccard descrita en el capítulo anterior. Esta se ha usado en las dos variantes comentadas con anterioridad; una en la que para cada imagen se haya la distancia entre las segmentaciones realizadas entre todos los pares posibles de observadores, hallando finalmente su promedio; la otra en la que se haya la cardinalidad de la intersección de todas las segmentaciones sobre una imagen, dividido la cardinalidad de la unión de las mismas. Estos cálculos fueron realizados tanto para la segmentación del núcleo como para la segmentación del citoplasma para cada célula procesada.

En la tabla 3.1, de modo resumido, se muestran los valores mínimos, máximos y promedio de la distancia de Jaccard empleando el primer método, sobre el conjunto de todas anotaciones sobre el conjunto imágenes, tanto para el núcleo, como para el citoplasma de las células. Como puede observarse de los valores medios (0.91 y 0.93), y teniendo en cuenta que esta distancia tiene valor máximo 1 (coincidencia total), existe un alto grado de coincidencia (por encima del 90 por ciento) para ambas partes de las células. Analizando los valores mínimos (0.83 y 0.89), nos percatamos de que no existe ningún caso crítico, puesto que aun estos valores podrían ser considerados satisfactorios. También debe observarse de que los valores son menores en el caso del núcleo comparado con el del citoplasma, debido a que de manera general los núcleos son más pequeños y una ligera desviación al segmentarlos repercute más en el valor calculado de la distancia de Jaccard.

Tabla 3.1. Distancia de Jaccard por pares de observadores sobre todas las imágenes del conjunto.

<b>Distancia de Jaccard</b> <b>Partes</b>	<b>Mínimo</b>	<b>Máximo</b>	<b>Media</b>
Núcleo	0.83	0.95	0.91
Citoplasma	0.89	0.97	0.93

De manera similar, en la tabla 3.2 se muestran los resultados relacionados con el cálculo de la distancia de Jaccard, solo que en este caso empleando el segundo método comentado. Se mantiene la misma observación realizada sobre las diferencias entre núcleo y citoplasma, al ser el primero de menor tamaño generalmente. Adicionalmente, como cabría esperar, puede notarse como en este caso los valores son visiblemente menores, lo cual es completamente lógico al involucrar en el cálculo todas las anotaciones sobre una misma imagen (cardinalidad de la intersección de todas las anotaciones sobre cardinalidad de la unión de todas las anotaciones sobre una misma imagen).

Tabla 3.2. Distancia de Jaccard calculada como la cardinalidad de la intersección dividido la cardinalidad de la unión para todas las anotaciones sobre una misma imagen.

<b>Distancia de Jaccard</b> <b>Partes</b>	<b>Mínimo</b>	<b>Máximo</b>	<b>Media</b>
Núcleo	0.42	0.84	0.73
Citoplasma	0.67	0.87	0.77

### 3.2 Comparación de algoritmos de segmentación empleando el índice de Williams.

A modo de ilustrar como puede ser utilizada la base de datos con las segmentaciones realizadas por los observadores, se han comparado los resultados de tres algoritmos de segmentación ampliamente extendidos en la literatura, haciendo uso de las mismas. Para ello se han empleado el índice de Williams y la distancia de Hausdorff, descritos en el capítulo anterior.

Los algoritmos para la segmentación automática de las imágenes fueron Otsu, Kmeans y Contornos Activos. Para el experimento fueron seleccionadas solo 10 imágenes. Esto se debe a que los algoritmos de segmentación fueron empleados en su forma más básica y solo se seleccionaron aquellas imágenes en las que estos se desempeñaban adecuadamente por simple inspección visual.

En la tabla 3.3 se muestran los resultados del cálculo del índice de Williams para los tres algoritmos de segmentación y cada una de las 10 imágenes empleadas en el experimento. Adicionalmente se muestran los límites inferior y superior respectivamente del intervalo de confianza para dicho índice. Recordemos que si el límite superior del intervalo de confianza es mayor que la unidad, podemos concluir que el resultado de la segmentación automática concuerda con los observadores así como estos concuerdan entre sí, o sea que el resultado es un miembro confiable del grupo. Analizando los resultados obtenidos, podemos concluir que para las imágenes seleccionadas el algoritmo de Otsu fue el más exitoso, con 7 imágenes de 10 segmentadas correctamente en correspondencia con el criterio de los observadores. Al mismo tiempo, el peor resultado lo obtuvo el algoritmo Kmeans con solo 3 imágenes segmentadas correctamente. Por último, el algoritmo de Contornos Activos dividió a partes iguales los aciertos y desaciertos.

Tabla 3.3. Índice de Williams utilizando la distancia de Hausdorff para los tres algoritmos de segmentación automáticos empleados

<b>Métodos de segmentación</b>	<b>I1</b>	<b>I2</b>	<b>I3</b>	<b>I4</b>	<b>I5</b>	<b>I6</b>	<b>I7</b>	<b>I8</b>	<b>I9</b>	<b>I10</b>	<b>Media</b>
Otsu	1.40 1.38-1.42	0.43 0.39-0.47	1.07 1.03-1.11	1.08 1.03-1.13	0.39 0.37- 0.41	1.13 1.09-1.18	0.84 0.81-0.87	1.15 1.13-1.17	1.17 1.14-1.20	1.10 1.07-1.14	0.97
Kmeans	1.08 1.05-1.12	0.52 0.47-0.57	0.50 0.47-0.53	0.16 0.15-0.17	0.35 0.34-0.37	0.47 0.46-0.49	0.64 0.62-0.67	0.85 0.82-0.88	1.15 1.12-1.19	0.97 0.94-1.01	0.66
Active	1.35 1.32-1.39	0.38 0.35- 0.41	1.15 1.12-1.18	0.51 0.49-0.54	1.02 1.00-1.05	1.21 1.18-1.24	0.67 0.64-0.71	0.45 0.44-0.47	1.03 1.00-1.06	0.50 0.48-0.53	0.82

## Conclusiones

En este trabajo se arribaron a las siguientes conclusiones:

- El procedimiento adoptado para la adquisición de las imágenes así como el algoritmo de corrección de campo plano (*flat field correction*) empleado en su rehabilitación garantizó la calidad de las mismas.
- El software ImageJ empleado en la anotación de las imágenes constituye una herramienta adecuada para tal fin, permitiendo un manejo fácil, flexible e intuitivo.
- El empleo de la distancia Jaccard en las dos formas propuestas para la medición de la variabilidad inter-observador arroja que existe una alta coincidencia entre las segmentaciones realizadas por los observadores.
- La base de datos anotada obtenida puede ser de gran utilidad para la prueba y puesta a punto de algoritmos como fue mostrado en las pruebas realizadas.

## Recomendaciones

- Se recomienda extender la base de datos con un mayor número de imágenes y que todos los tipos de leucocitos estén igualmente representados.
- Podrían incluirse más observadores y que cada uno haga varias anotaciones sobre una misma imagen lo cual permitiría hacer estudios intra-observador.

**REFERENCIAS BIBLIOGRÁFICAS**

- [1] Wikipedia. (2010), *Microscopio*. Disponible en: <http://es.wikipedia.org>, accedido el 13/4/2010.
- [2] Michael W. Davidson (1999), Mortimer Abramowitz, *Optical Microscopy*, *National High Magnetic Field Laboratory, The Florida State University*
- [3] Wikipedia. (2010), *Microscopio óptico*. Disponible en: <http://es.wikipedia.org>, accedido el 24/4/2010.
- [4] Brian L. Kuyatt, Robert Weaver, Philip Merlo (2005). *Image Capture Methods*.
- [5] María del Rocío Ochoa Montiel (2006). *Clasificación de Leucocitos utilizando Visión por Computadora*, Universidad Autónoma de Tlaxcala, Facultad de Ciencias Básicas, Ingeniería y Tecnología.
- [6] Bagby GC. Leukopenia and leukocytosis. In: Goldman L, Ausiello D, eds. *Cecil Medicine*. 23rd ed. Philadelphia, Pa: Saunders Elsevier; 2007:chap 173.
- [7] J. B. J. R. H. Vélez, W. Rojas, *Hematología*, Fundamentos de Medicina, 6th ed., Cuellar F. Falabella.
- [8] Oppenheim (1998). *Manual para técnicos de laboratorio*. México D.F.
- [9] J. H. I. Davidsohn (1991). *Diagnostico Clínico por el Laboratorio* (6th ed.). Salvat, Ed.
- [10] M. F. Escobar (1994). *Módulo de Correlación Clínica en Semiología General*, Universidad Autónoma de Manizales, Facultad de Odontología
- [11] Diana Lorena Toro B., Luis Eduardo Arcila B., Neil Guerrero (2007), *Reconocimiento y Conteo Automático de Glóbulos Blancos Utilizando MATLAB*, Universidad Autónoma de Manizales.
- [12] P. Besl and N. McKay (1992) *A method for registration of 3-D shapes*, IEEE Trans. Pattern Anal. Machine Intell., vol. 14, pp. 239–256.



- [13] John C. Russ (2007). *The Image Processing Handbook*, Materials science and Engineering Department, North Carolina State University, 5th ed. CRC Taylor & Francis Group, Ed
- [14] Wikipedia. (2010), *Cámara digital*. Disponible en: <http://es.wikipedia.org>, accedido el 30/10/2010.
- [15] Rafael González, Richard E. Woods (2001). *Digital Image Processing*. 2nd ed. Prentice Hall, Upper Saddle River, New Jersey.
- [16] María Gabriela Bustos Flores (2009). *Reconocimiento de células sanguíneas leucocitarias mediante redes neuronales*, Universidad Mayor de San Andrés.
- [17] Jounquera-Carneiro (1999). *Células de la sangre, Histología básica*, 6ta ed: Atlas, Cap.12.
- [18] Oppenheim (1998). *Manual para técnicos de laboratorio*. México D.F.
- [19] Jounqueira-Carneiro (1999). *Hematopoyético*, Histología Básica, 6ta ed: Atlas, Cap.13, pp.1-9.
- [20] E. A. Geiser, D. A. Conetta, M. C. Limacher, V. O. Stockton, L. H. Olivier, and B. Jones (1990). A second-generation computer-based edge detection algorithm for short-axis two-dimensional echocardiographic images: *Accuracy and improvement in interobserver variability*, J. Amer. Soc. Echocardiol., vol. 3, pp. 79-90.
- [21] C. deGraaf, A. Koster, K. Vincker, and M. Vierger (1992). *A methodology for the validation of images segmentation algorithms*, in Proc. IEEE Symp. Computer-base Medical Systems, pp. 17-24.
- [22] Chen M., Kanade, T., Pomerleau, D., & Rowely (1999), H.A. , *Anomaly Detection through registration*, Pattern Recognition, pp. 32, 113-128.
- [23] Vikram Chalana and Yongmin Kim (1997), A methodology for evaluation of boundary detection algorithms on medical images, *IEEE Transactions on Medical Imaging*, pp. 16(5):642-652.
- [24] L. H. Staib and J. S. Duncan (1989). *Left ventricular analysis from cardiac images using deformable models*, IEEE Comput. in Cardiol., Mag., pp. 427-430.

- [25] Cellmigration (2010). *Flat Field Correction*. Correction for Uneven Illumination. Disponible en: [http://www.cellmigration.org/resource/biosensors/ppt-pdf/2flat\\_field.pdf](http://www.cellmigration.org/resource/biosensors/ppt-pdf/2flat_field.pdf), accedido el 17/11/2010.
- [26] D. W. Davidson, C. Fröjdh, V. O'Shea, H-E. Nilsson, M. Rahman (2003). *Limitations to flat-field correction methods when using an X-ray spectrum*, Department of Physics and Astronomy, University of Glasgow, Glasgow G12
- [27] Michael W. Davidson (2002). *National High Magnetic Field Laboratory*, The Florida State University, Tallahassee, Florida, 32310.
- [28] J. J. Gerbrands (1988), *Segmentation of Noise Images*, Delft, The Netherlands.
- [29] Wikipedia. (2010), *ImageJ*. Disponible en: <http://es.wikipedia.org>, accedido el 7/12/2010.
- [30] Rasband, W.S. (1997-2004), ImageJ, National Institutes of Health, Bethesda, Maryland, USA, <http://rsb.info.nih.gov/ij/>.
- [31] David Thomas Gering (2003), *Recognizing deviations from normalcy for brain tumor segmentation*, Electrical Engineering and Computer Science. Massachusetts Institute of Technology.
- [32] Edward Ashton, Webster (2004), *Method and system for automatic identification and quantification of abnormal anatomical structures in medical images*, NY (US), pp 6.
- [33] E. A. Geiser, D. A. Conetta, M. C. Limacher, V. O. Stockton, L. H. Olivier, and B. Jones (1990). A second-generation computer-based edge detection algorithm for short-axis two-dimensional echocardiographic images: *Accuracy and improvement in interobserver variability*, J. Amer. Soc. Echocardiol., vol. 3, pp. 79-90.
- [34] A. Hammoude (2001), An empirical parameter selection method for endocardial border identification algorithms, *Computerized Medical Imaging and Graphics*, pp. 25:33-45.
- [35] Pascal Bamford (2003). *Empirical comparison of cell segmentation algorithms using an annotated dataset*, Department of information technology and Electrical Engineering The University of Queensland Brisbane, Australia.

- [36] P. Gil, F. Torres, F.G Ortiz (2004). *Detección de objetos por objetos de segmentación multinivel combinada de espacios de color*, Dto. De Física, Ingeniería de Sistemas y Teoría de la Señal, Universidad de Alicante.
- [37] J. B. Mena (2002), *Vectorización automática de una imagen binaria mediante K-means y degeneración de la triangulación de Delaunay*, Departamento de Matemáticas de la Universidad de Alcalá de Henares.
- [38] Alberto Cardona, Norberto Nigro, Victorio Sonzogni, Mario Storti (2006). *Segmentacion de imágenes digitales 3D basados en regiones y contornos activos para las generación de mallas de superficie*, Universidad Nacional del Centro de la Prov. de Buenos Aires.
- [39] Matias, J., Becilla, C. Chavez, P. (2009). *Segmentación de Imágenes medicas para detección de detalles*, Facultad de Ingeniería en Electricidad y Computación.
- [40] Roberto Carlos Chumbimuni Contreras, C. López and A. Bosnjak (2008). *Segmentación Utilizando Contornos Activos (Snakes) de Imágenes de Resonancia Magnética del Cerebro*, IFMBE Proceedings, 2008, Volume 18, Part 2, 313-316.

## ANEXOS

Los códigos de los programas en Matlab empleados en la realización de este trabajo se muestran a continuación:

### I. Función para el pre-procesamiento de las imágenes adquiridas. Corrección de campo plano (*flat field correction*).

```
function res = flat_field_correction(raw, background, dark)
raw = im2double(raw);
background = im2double(background);
dark = im2double(dark);
d = background - dark;
res = ((raw - dark) ./ d);
dR = d(:, :, 1);
dG = d(:, :, 2);
dB = d(:, :, 3);
res(:, :, 1) = res(:, :, 1) * mean(dR(:));
res(:, :, 2) = res(:, :, 2) * mean(dG(:));
res(:, :, 3) = res(:, :, 3) * mean(dB(:));
```

### II. Función para el cálculo de distancia de Hausdorff.

```
function hd = hausdorff_dist(P,Q)
% Calculates the Hausdorff Distance between P and Q%
% hd = HausdorffDist(P,Q)
% Calculates the Hausdorff Distance, hd, between two sets of points, P and Q
% (which could be two trajectories) in two dimensions. Sets P and Q must
% therefore be 2-dimensional matrices, though not necessarily of equal size.%
% The Directional Hausdorff Distance (dhd) is defined as:
% dhd(P,Q) = max p c P [ min q c Q [ ||p-q|| ] ].
% Intuitively dhd finds the point p from the set P that is farthest from any
% point in Q and measures the distance from p to its nearest neighbor in Q.
```

---

```

% The Hausdorff Distance is defined as  $\max\{dhd(P,Q),dhd(Q,P)\}$ 
% ZCD Oct 2009
sP = size(P); sQ = size(Q);
if ndims(P)>2 || ndims(Q)>2
    error('Inputs must be 2-D vectors')
elseif isempty(find(size(P)==2,1)) || isempty(find(size(Q)==2,1))
    error('Inputs must be 2-D vectors')
end
% reorient vectors
if sP(2)~=2, P=P'; end
if sQ(2)~=2, Q=Q'; end
% obtain all possible point comparisons
iP = repmat(1:length(P),[1,length(Q)]);
iQ = repmat(1:length(Q),[length(P),1]);
combos = [iP,iQ(:)];
% get distances for each point combination
cP=P(combos(:,1),:); cQ=Q(combos(:,2),:);
dists = sqrt( (cP(:,1)-cQ(:,1)).^2 + (cP(:,2)-cQ(:,2)).^2 );
% Now create a matrix of distances where D(n,m) is the distance of the nth
% point in P from the mth point in Q. The maximum distance from any point
% in Q from P will be max(D,[],1) and the maximum distance from any point
% in P from Q will be max(D,[],1);
D = reshape(dists,length(P),[]);
% Obtain the value of the point, p, in P with the largest minimum distance
% to any point in Q.
vp = max(min(D,[],2));
% Obtain the value of the point, q, in Q with the targets minimum distance
% to any point in P.
vq = max(min(D,[],1));
hd = max(vp,vq);

```

### III. Función para la segmentación empleando el método de Otsu.

```
function [IDX,sep] = otsu_seg(I,n)
%OTSU Global image thresholding/segmentation using Otsu's method.
%  IDX = OTSU(I,N) segments the image I into N classes by means of Otsu's
%  N-thresholding method. OTSU returns an array IDX containing the cluster
%  indices (from 1 to N) of each point. Zero values are assigned to
%  non-finite (NaN or Inf) pixels.
%  IDX = OTSU(I) uses two classes (N=2, default value).
%  [IDX,sep] = OTSU(...) also returns the value (sep) of the separability
%  criterion within the range [0 1]. Zero is obtained only with data
%  having less than N values, whereas one (optimal value) is obtained only
%  with N-valued arrays.
%  Notes:
%  -----
%  It should be noticed that the thresholds generally become less credible
%  as the number of classes (N) to be separated increases (see Otsu's
%  paper for more details).
%  If I is an RGB image, a Karhunen-Loeve transform is first performed on
%  the three R,G,B channels. The segmentation is then carried out on the
%  image component that contains most of the energy.
%  Example:
%  -----
%  load clown
%  subplot(221)
%  X = ind2rgb(X,map);
%  imshow(X)
%  title('Original','FontWeight','bold')
%  for n = 2:4
%    IDX = otsu(X,n);
%    subplot(2,2,n)
%    imagesc(IDX), axis image off
%    title(['n = ' int2str(n)],'FontWeight','bold')
```

```
% end
% colormap(gray)
% Reference:
% -----
% Otsu N, <a href="matlab:web('http://dx.doi.org/doi:10.1109/TSMC.1979.4310076')">A
Threshold Selection Method from Gray-Level Histograms</a>,
% IEEE Trans. Syst. Man Cybern. 9:62-66;1979
% See also GRAYTHRESH, IM2BW
% -- Damien Garcia -- 2007/08, revised 2010/03
% Visit my <a
% href="matlab:web('http://www.biomecardio.com/matlab/otsu.html')">website</a> for
more details about OTSU
error(nargchk(1,2,nargin))
% Check if is the input is an RGB image
isRGB = isrgb(I);
assert(isRGB | ndims(I)==2,...
    'The input must be a 2-D array or an RGB image.')
%% Checking n (number of classes)
if nargin==1
    n = 2;
elseif n==1;
    IDX = NaN(size(I));
sep = 0;
return
elseif n~=abs(round(n)) || n==0
    error('MATLAB:otsu:WrongNValue',...
        'n must be a strictly positive integer!')
elseif n>255
    n = 255;
    warning('MATLAB:otsu:TooHighN',...
        'n is too high. n value has been changed to 255.')
end
I = single(I);
```

---

```

%% Perform a KLT if isRGB, and keep the component of highest energy
if isRGB
    sizI = size(I);
    I = reshape(I,[],3);
    [V,D] = eig(cov(I));
    [tmp,c] = max(diag(D));
    I = reshape(I*V(:,c),sizI(1:2)); % component with the highest energy
end
% Convert to 256 levels
I = I-min(I(:));
I = round(I/max(I(:))*255);
%% Probability distribution
unI = sort(unique(I));
nbins = min(length(unI),256);
if nbins==n
    IDX = ones(size(I));
    for i = 1:n, IDX(I==unI(i)) = i; end
    sep = 1;
    return
elseif nbins<n
    IDX = NaN(size(I));
    sep = 0;
    return
elseif nbins<256
    [histo,pixval] = hist(I(:),unI);
else
    [histo,pixval] = hist(I(:),256);
end
P = histo/sum(histo);
clear unI
%% Zeroth- and first-order cumulative moments
w = cumsum(P);
mu = cumsum((1:nbins).*P);

```



```

%% Maximal sigmaB^2 and Segmented image
if n==2
    sigma2B =...
        (mu(end)*w(2:end-1)-mu(2:end-1)).^2./w(2:end-1)./(1-w(2:end-1));
    [maxsig,k] = max(sigma2B);
    % segmented image
    IDX = ones(size(I));
    IDX(I>pixval(k+1)) = 2;
    % separability criterion
    sep = maxsig/sum(((1:nbins)-mu(end)).^2.*P);
elseif n==3
    w0 = w;
    w2 = fliplr(cumsum(fliplr(P)));
    [w0,w2] = ndgrid(w0,w2);
    mu0 = mu./w;
    mu2 = fliplr(cumsum(fliplr((1:nbins).*P))./cumsum(fliplr(P)));
    [mu0,mu2] = ndgrid(mu0,mu2);

    w1 = 1-w0-w2;
    w1(w1<=0) = NaN;
    sigma2B =...
        w0.*(mu0-mu(end)).^2 + w2.*(mu2-mu(end)).^2 +...
        (w0.*(mu0-mu(end)) + w2.*(mu2-mu(end))).^2./w1;
    sigma2B(isnan(sigma2B)) = 0; % zeroing if k1 >= k2
    [maxsig,k] = max(sigma2B(:));
    [k1,k2] = ind2sub([nbins nbins],k);
    % segmented image
    IDX = ones(size(I))*3;
    IDX(I<=pixval(k1)) = 1;
    IDX(I>pixval(k1) & I<=pixval(k2)) = 2;
    % separability criterion
    sep = maxsig/sum(((1:nbins)-mu(end)).^2.*P);
else

```

---

```

    k0 = linspace(0,1,n+1); k0 = k0(2:n);
[k,y] = fminsearch(@sig_func,k0,optimset('TolX',1));
k = round(k*(nbins-1)+1);
    % segmented image
    IDX = ones(size(I))*n;
    IDX(I<=pixval(k(1))) = 1;
    for i = 1:n-2
        IDX(I>pixval(k(i)) & I<=pixval(k(i+1))) = i+1;
    end
    % separability criterion
sep = 1-y;
    end
    IDX(~isfinite(I)) = 0;
% Function to be minimized if n>=4
    function y = sig_func(k)
muT = sum((1:nbins).*P);
sigma2T = sum(((1:nbins)-muT).^2.*P);
    k = round(k*(nbins-1)+1);
k = sort(k);
    if any(k<1 | k>nbins), y = 1; return, end
k = [0 k nbins];
    sigma2B = 0;
    for j = 1:n
wj = sum(P(k(j)+1:k(j+1)));
        if wj==0, y = 1; return, end
muj = sum((k(j)+1:k(j+1)).*P(k(j)+1:k(j+1)))/wj;
        sigma2B = sigma2B + wj*(muj-muT)^2;
    end
    y = 1-sigma2B/sigma2T; % within the range [0 1]
    end
end
function isRGB = isrgb(A)
% --- Do we have an RGB image?

```

```
% RGB images can be only uint8, uint16, single, or double
isRGB = ndims(A)==3 && (isfloat(A) || isa(A,'uint8') || isa(A,'uint16'));
% ---- Adapted from the obsolete function ISRGB ----
if isRGB&&isfloat(A)
    % At first, just test a small chunk to get a possible quick negative
    mm = size(A,1);
    nn = size(A,2);
    chunk = A(1:min(mm,10),1:min(nn,10),:);
    isRGB = (min(chunk(:))>=0 && max(chunk(:)<=1);
% If the chunk is an RGB image, test the whole image
    if isRGB, isRGB = (min(A(:))>=0 && max(A(:)<=1); end
end
end
```

#### IV. Función para descartar objetos que tocan el borde en las imágenes segmentadas.

```
function imout = discard_objects_touching_border(im)
imout = im;
terminate = false;
while(~terminate)
    y1 = find(imout(1, :));
    if ~isempty(y1)
        imout(bwselect(imout, y1(1), 1, 8)) = 0;
    end
    y2 = find(imout(end, :));
    if ~isempty(y2)
        imout(bwselect(imout, y2(1), size(imout, 1), 8)) = 0;
    end
    x1 = find(imout(:, 1));
    if ~isempty(x1)
        imout(bwselect(imout, 1, x1(1), 8)) = 0;
    end
end
```

```

    x2 = find(imout(:, end));
    if ~isempty(x2)
    imout(bwselect(imout, size(imout, 2), x2(1), 8)) = 0;
    end
    if isempty(y1) && isempty(y2) && isempty(x1) && isempty(x2)
    terminate = true;
    end
end

```

#### **V. Función para descartar objetos pequeños y muy grandes en las imágenes segmentadas.**

```

function imout = discard_objects_outside_size(im, minA, maxA)
imout = im;
[L, n] = bwlabel(imout, 8);
s = regionprops(imout, L, 'area');
for i = 1 : n
    if s(i).Area < minA || s(i).Area > maxA
    imout(L == i) = 0;
    end
end

```

#### **VI. Script para el cálculo de la distancia de Jaccard por parejas de observadores.**

```

imdir = 'C:\Documents and Settings\afalcon.UCLV\Desktop\Tesis de Reynier\Imagenes
segmentadas a mano';
imext = 'tif';
imh = 256; imw = 256;
impart = 'n'; % or c
nobs = 20;
folders = dir(imdir);
nfolders = length(folders);
djaccard = [];
for i = 1 : nfolders

```

---

```

if folders(i).isdir && ~strcmp(folders(i).name, '.') && ~strcmp(folders(i).name, '..')
    disp(folders(i).name);
    impath = fullfile(imdir, folders(i).name);
    files = dir(fullfile(impath, ['*.', imext]));
    if isempty(files)
        impath = fullfile(imdir, folders(i).name, [folders(i).name, '_', impart]);
        files = dir(fullfile(impath, ['*.', imext]));
    elseif strcmp(impart, 'c')
        continue;
    end
    djs = 0;
    nfiles = length(files);
    for j = 1 : nfiles - 1
        im1 = imread(fullfile(impath, files(j).name)) > 1;
        for k = j + 1 : nfiles
            im2 = imread(fullfile(impath, files(k).name)) > 1;
            djs = djs + length(find(im1 & im2))/length(find(im1 | im2));
        end
    end
    djs = djs / (nfiles*(nfiles-1)/2);
    djaccard = [djaccard djs];
end
end

```

**VII. Cálculo de la distancia de Jaccard como la cardinalidad de la intersección dividida la cardinalidad de la unión de las segmentaciones de una imagen.**

```

imdir = 'C:\Documents and Settings\afalcon.UCLV\Desktop\Tesis de Reynier\Imagenes
segmentadas a mano';
imext = 'tif';
imh = 256; imw = 256;
impart = 'n'; % n (nucleo) o c (citoplasma)
nobs = 20;

```

```
folders = dir(imdir);
nfolders = length(folders);
djaccard = [];

for i = 1 : nfolders
    if folders(i)..isdir && ~strcmp(folders(i).name, '.') && ~strcmp(folders(i).name, '..')
        disp(folders(i).name);
        impath = fullfile(imdir, folders(i).name);
        files = dir(fullfile(impath, ['*.', imext]));
        if isempty(files)
            impath = fullfile(imdir, folders(i).name, [folders(i).name, '_', impart]);
            files = dir(fullfile(impath, ['*.', imext]));
        elseif strcmp(impart, 'c')
            continue;
        end
        imunion = zeros(imh, imw);
        iminter = ones(imh, imw);
        for j = 1 : length(files)
            im = imread(fullfile(impath, files(j).name));
            imunion = imunion | (im > 0);
            iminter = iminter & (im > 0);
        end
        djaccard = [djaccard length(find(iminter)) / length(find(imunion))];
    end
end
```

### VIII. Script para el cálculo del índice de Williams

```
seg_algorithms = {'otsu', 'kmeans', 'active'}; % los nombres deben coincidir con los de las carpetas
obsfolder = 'observadores';
imsegext = 'bmp';
imobsext = 'tif';
```

```

nseg_algorithms = length(seg_algorithms);

% folderobs = dir(obsfolder);
% nfolderobs = length(folderobs);
% Pnv = [];
% Pnh = [];
% c = 1;
% for i = 1 : nfolderobs
%         if folderobs(i).isdir    &&    ~strcmp(folderobs(i).name,  '.')    &&
~strcmp(folderobs(i).name, '..')
%         disp(folderobs(i).name)
%         filesobs = dir(fullfile(obsfolder, folderobs(i).name, ['*.', imobsext]));
%         nfilesobs = length(filesobs);
%         sdv = 0;
%         sdh = 0;
%         c1 = 1;
%         for j = 1 : nfilesobs-1
%             im1 = (imread(fullfile(obsfolder, folderobs(i).name, filesobs(j).name)) > 0) + 1;
%             [x, y] = find(edge(im1, 'sobel')); im11 = [x, y];
%             for k = j+1 : nfilesobs
%                 im2 = (imread(fullfile(obsfolder, folderobs(i).name, filesobs(k).name)) > 0) +
1;
%                 v = vinet(im1, im2);
%                 if v == 0
%                     disp([filesobs(j).name, ' ', filesobs(k).name])
%                 else
%                     sdv = sdv + 1/v;
%                     Dv(c, c1) = 1/v;
%                 end
%                 [x, y] = find(edge(im2, 'sobel')); im22 = [x, y];
%                 h = hausdorff_dist(im11, im22);
%                 if h == 0
%                     disp([filesobs(j).name, ' ', filesobs(k).name])

```

---

```

%         else
%             sdh = sdh + 1/h;
%             Dh(c, c1) = 1/h;
%             c1 = c1 + 1;
%         end
%     end
% end
%     Pnv(c) = sdv/((nfilesobs*(nfilesobs-1))/2)
%     Pnh(c) = sdh/((nfilesobs*(nfilesobs-1))/2)
%     Pnname{c} = folderobs(i).name;
%     c = c + 1;
% end
% end

Dcov = [];
Dcoh = [];
for i = 1 : nseg_algorithms
    filesseg = dir(fullfile(seg_algorithms{i}, ['*.', imsegeft]));
    nfilesseg = length(filesseg);
    c = 1;
    for j = 1 : nfilesseg
        ims = (imread(fullfile(seg_algorithms{i}, filesseg(j).name)) > 0) + 1;
        [x, y] = find(edge(ims, 'sobel')); im11 = [x, y];
        [PATHSTR, NAME] = fileparts(filesseg(j).name);
        filesobs = dir(fullfile(obsfolder, NAME, ['*.', imobseft]));
        nfilesobs = length(filesobs);
        sdv = 0;
        sdh = 0;
        for k = 1 : nfilesobs
            imo = (imread(fullfile(obsfolder, NAME, filesobs(k).name)) > 0) + 1;
            v = vinet(ims, imo);
            sdv = sdv + 1/v;
            [x, y] = find(edge(imo, 'sobel')); im22 = [x, y];

```



```

    h = hausdorff_dist(im11, im22);
    sdh = sdh + 1/h;
    Dcov(j, k, i) = 1/v;
    Dcoh(j, k, i) = 1/h;
    c = c + 1;
end
sdv = sdv/nfilesobs;
sdh = sdh/nfilesobs;
resvinet(i, j) = sdv/Pnv(j);
reshd(i, j) = sdh/Pnh(j);
filesseg(j).name
Pnname{j}
end
end

```

**IX. Script para el cálculo del intervalo de confianza del índice de Williams empleando el método de jackknife.**

```

P0h = [];
for i = 1 : size(Dcoh, 3)
    for j = 1 : size(Dcoh, 1)
        for k = 1 : size(Dcoh, 2)
            P0h(j, k, i) = (sum(Dcoh(j, :, i)) - Dcoh(j, k, i))/size(Dcoh, 2);
        end
    end
end
nobs = 20;
Pnh = [];
for i = 1 : size(Dh, 1)
    for l = 1 : nobs
        s = 0;
        c = 0;
        for j = 1 : nobs - 1

```

---

```

        for k = j + 1 : nobs
            c = c + 1;
            if l ~= j && l ~= k
                s = s + Dh(i, c);
            end
        end
    end
    Pnh(i, l) = s / ((nobs*(nobs-1))/2);
end
end
Ii(:, :, 1) = P0h(:, :, 1) ./ Pnh;
Ii(:, :, 2) = P0h(:, :, 2) ./ Pnh;
Ii(:, :, 3) = P0h(:, :, 3) ./ Pnh;
se = [];
CI = [];
for i = 1 : size(Ii, 3)
    for j = 1 : size(Ii, 1)
        Ip = 1/size(Ii, 2)*sum(Ii(j, :, i));
        se = sqrt((1/(size(Ii, 2)-1))*sum((Ii(j, :, i)-Ip).^2));
        CI(j, 1, i) = Ip - 1.96*se;
        CI(j, 2, i) = Ip + 1.96*se;
    end
end
end

```