

A heuristic model for determining the sperm motility grade by video

Diego Gárate, Rosario Medina, Filomén Incahuanaco, César Beltrán
Pontificia Universidad Católica del Perú
Grupo de Reconocimiento de Patrones e inteligencia Artificial Aplicada (GRPIAA)
Lima, Perú

Abstract—Sperm motility analysis is an important part in the fertility analysis and at the same time is a good example of the tracking multiple objects and video surveillance problem from a computational point of view. Currently the most used method for this analysis is the direct inspection which is inaccurate, subjective, a not repeatable procedure, and difficult to teach. This article attempts to overcome these barriers using computer vision techniques and also proposes a heuristic model based on movement direction and Euclidean distance to track sperm in videos obtained from an artificial sperm cells simulator.

Keywords—Sperm Motility, tracking, image processing, computer vision

I. INTRODUCCIÓN

Para la Organización Mundial de la Salud (OMS) y el Comité Internacional para el Monitoreo de Tecnologías de Reproducción Asistida (ICMART), la infertilidad es una enfermedad del sistema reproductivo definida como la incapacidad de concebir un embarazo clínico después de 12 meses o más de intentarlo sin métodos de planificación [1].

Diversos estudios fueron realizados para determinar las causas que provocan la infertilidad en los hombres, entre ellas tenemos diferentes factores y enfermedades como: bajo volumen seminal, baja calidad espermática, problemas eyaculatorios, niveles bajos de Gonadotropina, síndrome adrenogenital, problemas tiroideos, necropermia, elevada densidad espermática, elevada viscosidad en el líquido seminal; entre otras [2].

En las últimas décadas, diferentes reportes sugieren que la calidad del semen en los hombres se viene reduciendo con el pasar de los años [3, 4]. Auger en [5], confirma estos estudios, obteniendo como resultado la disminución del conteo de espermatozoides (2.1% anual) y una reducción de la motilidad espermática (0.6% anual), a nivel mundial. Es así, que el análisis de la motilidad espermática se torna una parte importante en el análisis de calidad del semen [6].

Actualmente, la forma más usada para realizar este análisis es el método de inspección directa basado en observaciones de las muestras espermáticas a través del microscopio, este método es bastante subjetivo, pues los resultados varían de un especialista a otro [7]. Según Katz, estos tipos de procedimientos son inexactos, subjetivos, no repetibles y difíciles de enseñar [8].

La principal contribución de este trabajo es la propuesta de un modelo heurístico para determinar la dirección de movimiento de células espermáticas que permita mejorar el cálculo de su trayectoria en secuencias de video, a fin de determinar el grado de motilidad para cada una de ellas. Con este fin, fue implementado un simulador de células espermáticas y los resultados son analizados en base a pruebas estadísticas.

Siguiendo esta breve introducción, en la Sección II, se describe la implementación del simulador de células espermáticas artificiales. La Sección III, describe el método de detección de cabezas de células espermáticas, debido a que el análisis de motilidad debe realizarse a partir de la cabeza de la célula [6]. El modelo heurístico propuesto para determinar la dirección del movimiento y para construir la trayectoria de cada célula espermática en el video a analizar, es descrito en la Sección IV. Luego, en la Sección V, es realizada la clasificación de la motilidad en cuatro grados; además se realizó la comparación entre nuestra propuesta y el algoritmo de *Optical Flow*, en términos de eficacia. Finalmente, las conclusiones y trabajos futuros son descritos en la Sección VI.

II. SIMULADOR DE CÉLULAS ESPERMÁTICAS

Mediante la observación de 15 videos obtenidos bajo el microscopio con un lente de 40x de aumento, se determinó empíricamente que los espermatozoides en el líquido seminal presentan movimientos particulares, los cuales se pueden clasificar en tres categorías. A continuación, se definen las ecuaciones matemáticas de las funciones de movimiento de cada uno de ellas.

Movimiento Lineal:

$$(X_t, Y_t) = (X_0 + V_x * t, Y_0 + V_y * t) \quad (1)$$

Movimiento Sinusoidal:

$$(X_t, Y_t) = (X_0 + V_x * t, Y_0 + \sin(X_t * f) * A) \quad (2)$$

Movimiento Circular:

$$(X_t, Y_t) = (X_0 + \cos(V_x * t * f) * A, Y_0 + \sin(V_y * t * f) * A) \quad (3)$$

donde (X_t, Y_t) , es la posición del espermatozoide en un tiempo "t"; X_0 e Y_0 , se refieren a la posición inicial en la que aparecen las células espermáticas; "A", representa la amplitud de los movimientos oscilatorios y "f", indica la frecuencia de onda.

Para implementar este simulador se utilizó D3.js, la cual es una librería gratuita y de código abierto que permite la manipulación de datos con la finalidad de mejorar la visualización de la información [9]. En la Fig. 1, podemos ver una captura del simulador implementado, con un total de 7 células espermáticas, donde el número de espermatozoides por cada tipo de movimiento es un parámetro ingresado por el usuario. Las cabezas fueron modeladas como elipses cuyo segmento mayor es de 6.5 píxeles y 4 píxeles de segmento menor. La pieza media es una recta cuyo ancho es de 4 píxeles y la cola tiene un ancho de 2 píxeles y una longitud de 12 píxeles.



Figura 1: Imagen correspondiente al simulador de células espermáticas.

III. DETECCIÓN DE CÉLULAS ESPERMÁTICAS

En esta sección, se describe el método de detección de células espermáticas, el cual es un paso crítico en nuestra propuesta dado que, el seguimiento (*tracking*) será realizado a los puntos detectados durante este proceso. Primero, será descrito el algoritmo de pre-procesamiento de las imágenes. Luego, presentamos la función para la detección de contornos y el proceso de obtención de centros de masa de las células espermáticas.

A. Pre-Procesamiento

Para conseguir una imagen libre de ruidos, primero se convierte la imagen a escala de grises y luego aplicando un *threshold* con valor 127, se obtiene una imagen binaria. Posteriormente, se aplica el operador morfológico de erosión, para eliminar las colas y piezas medias de la imagen. En la Fig. 2, se puede observar el diagrama de flujo seguido en la fase de pre-procesamiento.

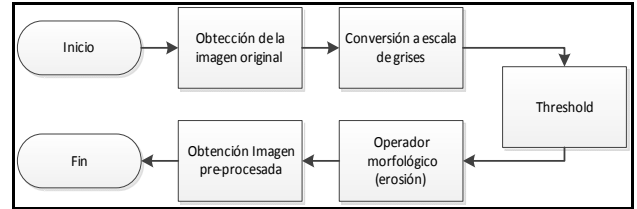


Figura 2: Diagrama de flujo de la fase de pre-procesamiento.

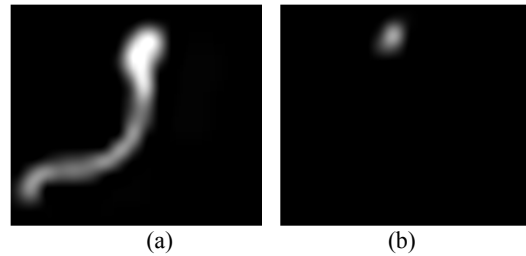


Figura 3: Aplicación del algoritmo de pre-procesamiento: (a) Imagen original y (b) Imagen obtenida como resultado

B. Detección de contorno y centros de masa

Este proceso consiste en obtener puntos característicos que representen una célula espermática. Con este fin, detectamos los contornos de las células espermáticas y obtenemos sus centros de masa a partir de las imágenes obtenidas luego del proceso de pre-procesamiento, como se puede ver en la Fig. 3(b). Utilizamos la función *findCounters* de OpenCV, la cual implementa el algoritmo propuesto por Suzuki y Be [10].

El centro de masa es un punto fijo representativo en el estudio de la geometría. Para una figura geométrica, este punto puede ser calculado usando el teorema de Green [11], como se detalla a continuación:

Dado $M = \iint \sigma(x, y) \partial A$; donde M representa la masa de la figura geométrica, su centro de masa es representado por:

$$\bar{X} = \frac{\iint x \sigma(x, y) \partial A}{M} \quad (4)$$

$$\bar{Y} = \frac{\iint y \sigma(x, y) \partial A}{M} \quad (5)$$

Estas fórmulas pueden ser aplicadas en imágenes como:

$$m_{ij} = \iint I(x, y) x^i y^j dx dy \quad (6)$$

donde $I(x, y)$ representa el valor de la imagen en el punto (x, y) . Luego, usando las fórmulas presentadas anteriormente podemos calcular el centro de masa en imágenes[12], como se muestra a continuación:

$$\bar{X} = \frac{m_{10}}{m_{00}}, \bar{Y} = \frac{m_{01}}{m_{00}} \quad (7)$$

Finalmente, se aplicó el método descrito para obtener los centros de masa de cada contorno detectado. Como se puede observar en la Fig. 4 (b) el centro de masa de cada célula espermática es representado por un círculo con radio de 5 píxeles para fines de visualización.

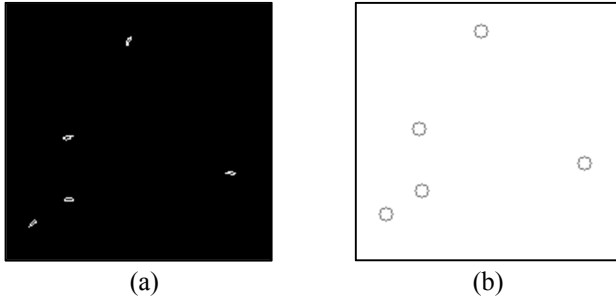


Figura 4: Resultados de la aplicación de nuestro método: (a) Detección de Contornos y (b) Centros de Masa

IV. MODELO HEURÍSTICO

En esta sección, serán descritos los detalles del modelo heurístico basado en dirección de movimiento y en la distancia Euclidiana. Primero, presentamos la función heurística del algoritmo y luego el resultado obtenido de la aplicación del modelo heurístico al problema de seguimiento de células espermáticas.

A. Función heurística

Para definir esta función nos basamos en la dirección de movimiento y la distancia Euclidiana, para ello utilizamos las siguientes funciones:

- Distancia Euclidiana:

$$d(A, B) = \sqrt{(Bx - Ax)^2 + (By - Ay)^2} \quad (8)$$

- Ángulo de movimiento:

$$\theta(\overrightarrow{AB}, \overrightarrow{CD}) = \cos^{-1}\left(\frac{ABx \cdot CDx + AB_y \cdot CD_y}{\sqrt{ABx^2 + AB_y^2} \cdot \sqrt{CDx^2 + CD_y^2}}\right) \quad (9)$$

donde A y B son puntos en el espacio de *frames* consecutivos. \overrightarrow{AB} , es un vector que representa el último tramo de la trayectoria encontrada hasta el momento de cada célula espermática y \overrightarrow{CD} es el vector entre el último punto de la trayectoria del espermatozoide a analizar y el punto que se intenta incorporar a la solución.

B. Aplicación al tracking de células espermáticas

La estructura de la solución está definida por una secuencia de puntos en el espacio que representan la trayectoria de una célula espermática.

$$Trayectoria(E_x) = \{P_0, P_1, P_2, \dots, P_n\}$$

donde E_x representa una célula espermática y P_i ($i = 0, 1, \dots, n$) son puntos en el plano.

A continuación se describe el conjunto de pasos de la heurística definida para el problema del seguimiento de células espermáticas.

- Sea E_t un conjunto de puntos en el plano, en un tiempo "t". Donde cada punto representa una posición en el plano de cada espermatozoide a seguir;
- Luego, se eligen dos subconjuntos $E_t(\text{nextPoint})$ y $E_{t-1}(\text{prevPoint})$ para todo $t > 0$, en el video analizado.
- Finalmente, la solución se obtiene al encontrar un valor mínimo que satisfaga la función heurística (10) para cada punto en el conjunto *nextPoint*, asociado a un punto en el conjunto *prevPoint*.

$$h(x) = f(x) + g(x) \quad (10)$$

donde:

- $f(x)$, representa a la distancia Euclidiana (8) y
- $g(x)$, describe la función ángulo (9)

V. ESTIMACIÓN DEL GRADO DE MOTILIDAD ESPERMÁTICA

La Organización Mundial de la Salud define cuatro grados de motilidad, en los cuales el movimiento de las células espermáticas puede ser clasificado [6]:

- Grado A o 1: espermatozoides con un movimiento progresivo, veloz, fuerte y en línea recta.
- Grado B o 2: espermatozoides con un movimiento no lineal, tienden a ir hacia delante pero con una trayectoria curvilínea.
- Grado C o 3: espermatozoides con un movimiento no progresivo, se observa movimiento pero sin desplazamiento.
- Grado D o 4: espermatozoides sin movimiento alguno (espermatozoides muertos).

Para poder clasificar las células espermáticas en los diferentes grados de motilidad, calculamos los siguientes valores: VSL, VCL y LIN [13].

- VSL: velocidad en línea recta, determinada por la distancia entre el punto inicial y el punto final de la trayectoria analizada entre el tiempo transcurrido.

$$VSL(Trayectoria(E_x)) = d(P_n, P_0) / t \quad (11)$$

- VCL: velocidad a lo largo de la curva, determinada por la distancia recorrida a lo largo de la trayectoria analizada entre el tiempo transcurrido.

$$VCL(Trayectoria(E_x)) = \frac{1}{t} * \sum_{t=0}^{t=n-1} d(P_t, P_{t+1}) \quad (12)$$

- LIN: ratio determinado por la división de VCL entre VSL

$$LIN(Trayectoria(E_x)) = \frac{VSL}{VCL} \quad (13)$$

Los valores VSL y LIN se utilizaron según lo descrito en la revista libre de Andrología para realizar la clasificación en clases de motilidad [13]. Los valores para cada clase en micrómetros por segundo ($\mu\text{m/s}$) se presentan a seguir:

$$\text{Grado 1} \begin{cases} VSL \geq 23\mu\text{m/s} \\ LIN \geq 0.58; \end{cases}$$

$$\text{Grado 2} \begin{cases} 23\mu\text{m/s} > VSL > 10\mu\text{m/s} \\ LIN \geq 0.58; \end{cases}$$

$$\text{Grado 3} \begin{cases} VSL > 10\mu\text{m/s} \\ LIN < 0.58; \end{cases}$$

$$\text{Grado 4} \{VSL \leq 10\mu\text{m/s}$$

VI. EVALUACIÓN DEL MODELO HEURÍSTICO

En esta sección serán descritos los resultados obtenidos por el modelo heurístico propuesto y la comparación de los mismos con el algoritmo Optical Flow. Con este fin, utilizamos 30 muestras obtenidas por el simulador implementado. Finalmente, se detallan la técnica utilizadas para la comparación de los resultados.

A. Resultados

En la Fig. 6 y 7, podemos visualizar las trayectorias encontradas por el algoritmo Optical Flow y el modelo heurístico propuesto, respectivamente.

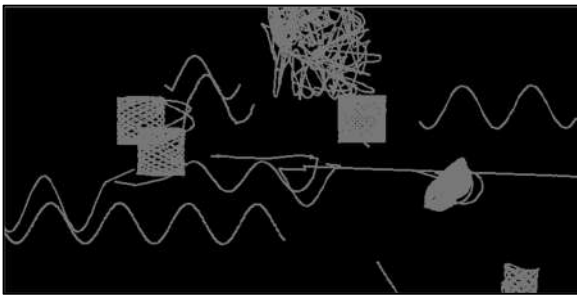


Figura 5: Trayectorias halladas mediante el algoritmo de Optical Flow.

El algoritmo de *Optical Flow* (ver Fig. 6), genera trayectorias erráticas en la parte central y superior de la imagen, estas trayectorias no corresponden a ninguna función de movimiento implementada en el simulador. Por el contrario en la Fig. 7 (modelo heurístico), se observan trayectorias semejantes a los patrones de movimiento implementados en el simulador.

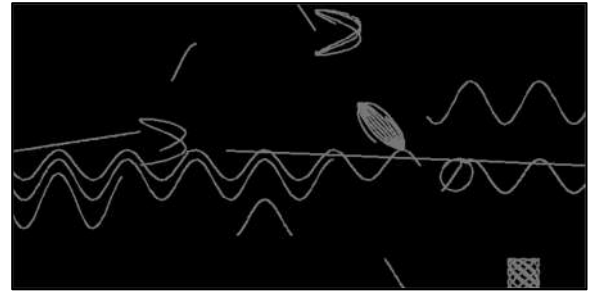


Figura 6: Trayectorias halladas mediante el modelo heurístico.

B. Comparación de Resultados

Para realizar la comparación se procedió a agrupar los resultados según el grado de motilidad hallado para cada célula espermática en la muestra, obteniendo así los porcentajes de cada clase como se muestra en la Tabla 1. Por ejemplo, en la Tabla 1 se observa que según el Optical Flow no existen células espermáticas de Grado 2, mientras que según la heurística propuesta hay 6.7% de dicho Grado, siendo el valor real de 13.3% dado por el simulador.

Tabla 1: Porcentajes según grado de motilidad para una muestra con un parámetro de 30 células espermáticas.

Grado de Motilidad	Porcentaje Simulador	Porcentaje Optical Flow	Porcentaje Modelo Heurístico
Grado 1	30.0%	36.7%	40.0%
Grado 2	13.3%	0.0%	6.7%
Grado 3	16.7%	20.0%	10.0%
Grado 4	40.0%	43.3%	43.3%

Una vez obtenidos los porcentajes, según el grado de motilidad para cada célula, se aplicó la siguiente función:

$$\text{Variación} = \left(\frac{1}{2m}\right) * \sum_{i=0}^{i=m} (h(i) - y(i))^2 \quad (14)$$

donde, m es la cantidad de grados distintos en este caso cuatro; $h(i)$ es el porcentaje obtenido para el grado i según el algoritmo aplicado; $y(i)$ es el porcentaje obtenido para el grado i según el simulador.

A menor valor de *Variación* más eficaz es el algoritmo, siendo el óptimo un valor de 0. Una vez hallado el valor de la variación para cada una de las 30 muestras tanto para el algoritmo *Optical Flow* como para el modelo heurístico propuesto, se realizó la comparación entre ambos algoritmos.

Debido a que el número de muestras no es mayor a treinta y la varianza poblacional es desconocida se determinó utilizar la prueba estadística *t-student* para la comparación de medias. La prueba requiere demostrar previamente que las varianzas muestrales son iguales y que las muestras siguen una distribución normal. Para ello, realizamos una prueba estadística de análisis de varianza conocida como prueba ANOVA [14] y la prueba de normalidad de Kolmogorov [15].

En la Tabla 2 se muestra el resultado de la prueba de Kolmogorov-Smirnov realizada en el software estadístico SPSS de IBM con la finalidad de demostrar la normalidad de ambas muestras.

Tabla 2: Prueba de Normalidad Kolmogorov-Smirnov

Resumen de la prueba de Hipótesis			
Hipótesis Nula	Prueba	Sig.	Resultado
La distribución de Optical Flow es normal con media 110.121 y desviación estándar 62.02	Prueba simple de Kolmogorov-Smirnov	0.2	Se retiene la hipótesis nula
La distribución de la Heurística es normal con media 74.453 y desviación estándar 47.16	Prueba simple de Kolmogorov-Smirnov	0.2	Se retiene la hipótesis nula

En las Tablas 3 y 4, se observan los resultados de realizar la prueba de ANOVA para determinar la igualdad de la varianza para el algoritmo de Optical Flow y la heurística propuesta respectivamente.

Por último comprobada la normalidad y la igualdad de varianzas muestrales en ambas muestras se procedió a realizar la prueba estadística t-student, la que permite determinar si la media poblacional del modelo heurístico es menor o igual a la media poblacional del algoritmo Optical Flow, como se puede observar en la Fig. 11 la hipótesis nula ha sido aceptada.

Tabla 3 Prueba ANOVA del modelo Optical Flow

Prueba Homogeneidad de las varianzas Optical Flow

Estadístico de Levene	dif1	dif2	Sig.
0.02	1	28	0.889

ANOVA

Optical Flow					
	Suma de Cuadrados	df	Promedio de los Cuadrados	F	Sig.
Entre los Grupos	252.184	1	252.184	0.063	0.803
En el Grupo	111305.544	28	3975.198		
Total	111557.728	29			

Tabla 4 Prueba ANOVA del modelo heurístico

Prueba de Homogeneidad de las varianzas Heurística

Estadístico de Levene	dif1	dif2	Sig.
0.058	1	28	0.812

ANOVA

Heurística					
	Suma de Cuadrados	df	Promedio de los Cuadrados	F	Sig.
Entre los Grupos	995.558	1	995.558	0.439	0.513
En el Grupo	63506.696	28	2268.096		
Total	64502.254	29			

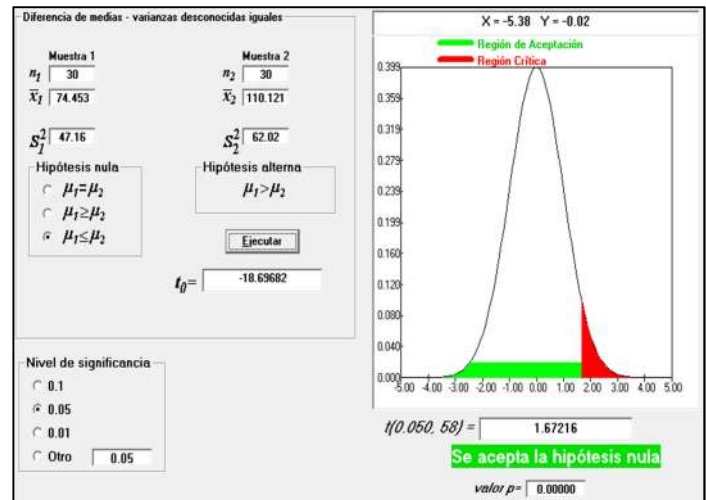


Figura 7: Prueba del estadístico t-student para comparación de medias poblacionales.

VII. CONCLUSIONES

Como resultado de la presente investigación se obtuvo un simulador de células espermáticas artificiales, el cual simula tres tipos de movimientos: movimiento lineal, movimiento sinusoidal y movimiento circular. Este simulador permite obtener videos con información objetiva acerca del movimiento realizado por las células artificiales con lo cual fue posible evaluar el modelo heurístico propuesto y el algoritmo de Optical Flow.

En cuanto a la detección de espermatozoides se obtuvo un algoritmo de pre-procesamiento capaz de obtener los centros de masa característicos de cada célula artificial y además se logró superar oclusiones menores es decir, al aplicar el algoritmo de detección si existieran células artificiales con alguna ligera superposición, se logra separar dichas células y obtener un punto característico para cada una.

Se obtuvo un modelo heurístico basado en dirección de movimiento el cual es capaz de determinar la trayectoria que sigue un espermatozoide tomando en consideración no solo la cercanía de los puntos sino también la dirección en la que se encuentra moviendo dicha célula. Se logró además implementar dicho modelo haciendo uso de la librería OpenCV y fue evaluado en términos de eficacia.

Por último, se realizó un análisis comparativo de medias cuyo resultado indicó que el modelo heurístico propuesto en este trabajo, aplicado al seguimiento de espermatozoides, es más eficaz que aplicar el algoritmo de Optical Flow sobre el mismo conjunto de células espermáticas.

VIII. TRABAJOS FUTUROS

En cuanto al simulador de células espermáticas artificiales, para el presente proyecto se enfocó en el movimiento de las células, se recomienda a modo de trabajo futuro estudiar más exhaustivamente el modelado de la pieza media y la cola de las células espermáticas artificiales, asimismo se puede modificar el simulador para generar información no solo de motilidad sino también de morfología, de tal forma que sea un simulador fiel a la realidad.

Sobre la detección de células espermáticas, se recomienda realizar pruebas en muestras reales de video obtenidas mediante microscopía a las cuales se les haya realizado un proceso de segmentación previo. Asimismo se recomienda realizar un estudio en términos de eficiencia del algoritmo de detección de espermatozoides.

Con respecto al modelo heurístico basado en dirección de movimiento y su aplicación, se recomienda incluir pesos en cada uno de los factores contemplados para la función heurística. Asimismo realizar un estudio comparativo con otras implementaciones de Optical Flow además de otros algoritmos que resuelvan la problemática del seguimiento a células espermáticas.

Por último en cuanto al análisis comparativo realizado, este se puede reforzar estudiando muestras que contengan: mayor cantidad de espermatozoides por video, menor cantidad de espermatozoides por video, disminuir o aumentar la calidad del video entre otros factores que puedan afectar el resultado del modelo.

REFERENCIAS

- [1] F. Zegers-Hochschild, G. D. Adamson, J. de Mouzon, O. Ishihara, R. Mansour, K. Nygren, *et al.*, "International Committee for Monitoring Assisted Reproductive Technology (ICMART) and the World Health Organization (WHO) revised glossary of ART terminology, 2009," *Fertil Steril*, vol. 92, pp. 1520-4, Nov 2009.
- [2] L. Dubin and R. D. Amelar, "Etiologic factors in 1294 consecutive cases of male infertility," *Fertility and Sterility*, vol. 22, pp. 469-474, August 1971.
- [3] D. S. Irvine, "Falling sperm quality," *BMJ*, vol. 309, 1994.
- [4] E. Carlsen, A. Giwercman, N. Keiding, and N. E. Skakkebaek, "Evidence for decreasing quality of semen during past 50 years," *Bmj*, vol. 305, pp. 609-13, Sep 12 1992.
- [5] J. Auger, J. M. Kunstmann, F. Czyglik, and P. Jouannet, "Decline in Semen Quality among Fertile Men in Paris during the Past 20 Years," *New England Journal of Medicine*, vol. 332, pp. 281-285, 1995.
- [6] O. L. Savkay and M. E. Yalcin, "Analysis of sperm motility with CNN architecture," in *Cellular Nanoscale Networks and Their Applications (CNNA), 2012 13th International Workshop on*, 2012, pp. 1-4.
- [7] H. Carrillo, J. Villarreal, M. Sotaquira, M. A. Goelkel, and R. Gutierrez, "A Computer Aided Tool for the Assessment of Human Sperm Morphology," in *Bioinformatics and Bioengineering, 2007. BIBE 2007. Proceedings of the 7th IEEE International Conference on*, 2007, pp. 1152-1157.
- [8] D. F. Katz, J. W. Overstreet, S. J. Samuels, P. W. Niswander, T. D. Bloom, and E. L. Lewis, "Morphometric Analysis of Spermatozoa in the Assessment of Human Male Fertility," *Journal of Andrology*, vol. 7, pp. 203-210, 1986.
- [9] M. Bostock. (2014). *d3js.org*. Available: www.d3js.org
- [10] S. Suzuki and K. be, "Topological structural analysis of digitized binary images by border following," *Computer Vision, Graphics, and Image Processing*, vol. 30, pp. 32-46, 4// 1985.
- [11] F. Chaumette, "Image moments: a general and useful set of features for visual servoing," *Robotics, IEEE Transactions on*, vol. 20, pp. 713-723, 2004.
- [12] OpenCV. (2014, 10/11/2014). *Structural Analysis and Shape Descriptors*.
- [13] N. I. J. Elia, M. Delfino, R. Mazzilli, T. Rossi and F. Mazzilli, "The Importance of the Sperm Motility Classes - Future Directions," *The Open Andrology Journal*, vol. 2, 2010.
- [14] M. B. Brown and A. B. Forsythe, "Robust Tests for the Equality of Variances," *Journal of the American Statistical Association*, vol. 69, pp. 364-367, 1974/06/01 1974.
- [15] H. W. Lilliefors, "On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown," *Journal of the American Statistical Association*, vol. 62, pp. 399-402, 1967/06/01 1967.