

Sistema de Reconocimiento de Posturas del Cuerpo Humano

Tatiana Aparicio Vergara y Cristóbal Sebastián González Dixon
tatiaparcio@gmail.com, sebasdixon@gmail.com

Carrera de Ingeniería Electrónica, Pontificia Universidad Javeriana. Bogota-Colombia, Noviembre de 2005.

RESUMEN

En este artículo se presenta el desarrollo de un sistema de visión artificial para el reconocimiento de posturas de las extremidades superiores del cuerpo humano. Siendo un sistema apto para trabajar con cualquier persona, sin importar su edad, raza, sexo o contextura física. Presenta facilidad en la adecuación del espacio y adaptabilidad ante en el cambio de ubicación de la región de interés en la imagen.

Palabras clave: Visión Artificial, reconocimiento de posturas, esqueletización, goniometría.

ABSTRACT

This paper presents the development of an artificial vision system for posture recognition of the upper extremities of human body. Being a system able to work with any person, doesn't matter it age, sex, race, or fisic. The system presents an easy way to enjance the working space and it is adaptable to the location change of the region of interest.

Key Words: Artificial Vision, Posture recognition, skeletonization, goniometric.

I. INTRODUCCION

El desarrollo de herramientas de comunicación hombre – máquina por medio de sistemas de visión artificial es cada vez mas popular, pues con estas se logra implementar métodos no intrusivos para el desarrollo de espacios interactivos en los que una persona puede realizar las actividades propuestas por el entorno virtual [BRC99] [CL04] [NKS03], por medio de comandos de control asociados al movimiento de su cuerpo. En este artículo se desarrolla un método de reconocimiento de posturas del cuerpo, con el cual es posible hacer el seguimiento de las extremidades superiores del cuerpo en tiempo real, para determinar los ángulos que forman las articulaciones de hombros y codos en imágenes capturadas por medio de video. En el capítulo 2 se hace un recuento de cada uno las pasos necesarios para obtener el resultado, proponiendo un nuevo método de esqueletización y una metodología de descripción goniométrica. En el capítulo 3 se muestran los resultados obtenidos terminando con una seria de conclusiones que muestran la importancia del desarrollo de este tipo de sistema.

II. METODOLOGÍA

El sistema implementado parte de la captura de la imagen por medio de una cámara web bajo condiciones de fondo e iluminación controladas, las cuales son esenciales para un buen resultado.

En la figura 1 se muestra el diagrama de flujo detallado del sistema propuesto, donde lo primero es almacenar una imagen del escenario vacío, para que durante el reconocimiento los píxeles diferentes sean solamente los que conforman la figura del usuario, luego si se hace la captura de la imagen, donde se extrae la silueta de la persona de tal forma que los píxeles que son fondo tomen el color negro y los que hacen parte de la persona el color blanco.

Luego, se realiza un preprocesamiento en el cual se realiza el realce y limpieza de la región de interés, en cada imagen capturada, teniendo en cuenta la eliminación de ruido y el filtrado de la imagen para tener una silueta uniforme.

Teniendo la silueta se procede a hacer una primera validación que permite que el proceso continúe o por el contrario se trunque para obtener una nueva captura lo mas pronto posible. Se verifica cuantos contornos hay en la imagen binaria de la silueta de tal forma que si detecta más de 4 asume que hay un objeto intruso en la escena, por lo que vuelve a capturar otra imagen y no permite continuar el algoritmo de este punto en adelante mientras la situación no se corrija. Si por el contrario si hay 4 o menos contornos se continúa haciendo la esqueletización.

La esqueletización consiste en encontrar una línea representativa de la silueta. Esto se logra por medio de adelgazamientos sucesivos, los cuales borran solamente píxeles del contorno en cada iteración. Para este adelgazamiento o *thinning* se debe cumplir con una serie de condiciones propuestas.

Con el esqueleto listo, se busca información relevante que permite eliminar información redundante y ruidosa. Para esto se determina la región del pecho. Este es un punto de cruce dentro del esqueleto que en teoría tiene cuatro ramificaciones: cabeza tronco y los dos brazos.

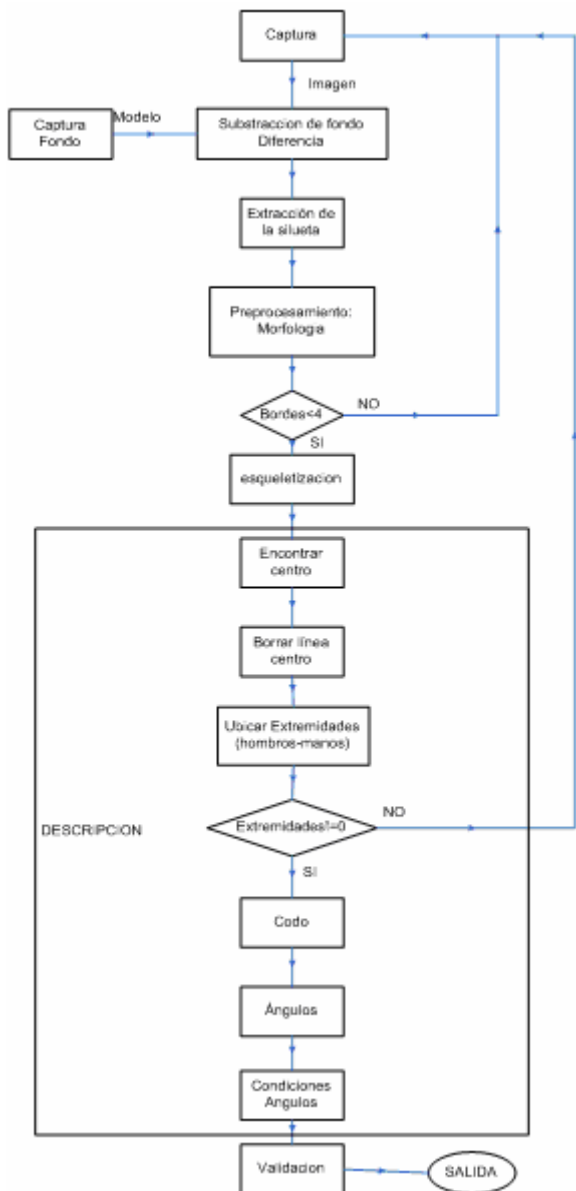


Figura 1. Diagrama de flujo

El proceso continúa eliminando la línea del tronco, que teniendo en cuenta la correcta postura del usuario es vertical y pasa por el punto del pecho, dejando así solo los píxeles que conforman los brazos, representados por dos líneas no conexas entre sí, donde sus extremos son la ubicación de los hombros y de las manos. La siguiente tarea es precisamente ordenar dichos extremos para clasificar: mano izquierda, mano derecha, hombro izquierdo y hombro derecho.

Con estos puntos organizados se hace un recorrido por cada una de las líneas para obtener datos como longitud y código de cadena, y encontrar los codos por características de las proporciones físicas del cuerpo humano las cuales fueron proporcionadas por Leonardo Davinci en su estudio de los cánones de las proporciones divinas, plasmados en su obra sobre el “Hombre Vitruviano”.

La tarea de reconocimiento continúa con el cálculo trigonométrico de los ángulos de los hombros y de los codos con respecto al eje vertical como se muestra en la

figura 2. A cada postura le corresponde un rango de estos ángulos con un margen de error para contrarrestar las diferencias que hay entre personas por su contextura física y por su concepción de las posturas propuestas. Por último se retorna cual es la postura.

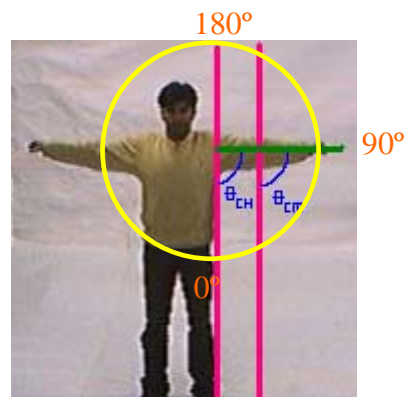


Figura 2. Ángulos codo y hombro

2.1 Adecuación del Entorno

Para el funcionamiento del sistema hay que tener en cuenta una serie de condiciones ambientales, las cuales deben ser controladas para asegurar que la escena no sufrirá cambios durante la ejecución de la actividad para la que sea implementado el sistema.

2.1.1 Luces

Para obtener una iluminación que proporcione óptimos resultados es suficiente con disponer lámparas de luz blanca, de forma que se ilumine todo el espacio de forma homogénea. Es conveniente el uso de luces omnidireccionales, con el fin de evitar focos de luz y regiones de sombra. No resulta conveniente usar luces muy fuertes, pues estas tienden a realzar las texturas presentes en la escena y ocasionar fuertes sombras. Un excelente resultado se puede obtener usando un par de lámparas fluorescentes tubulares, las cuales además de no disipar energía en forma de calor, ofrecen mas comodidad al usuario del sistema. Se sugiere ubicar las luces entre la persona y el fondo del escenario, para así tener mayor contraste en la imagen, pues con esto se logra iluminar el fondo, que refleja la luz y deja en contraluz a la persona, evitando confusión de colores y concentrándose en luz y sombra.

2.1.2 Fondo y entorno

Por su parte, es importante disponer de un espacio con dimensiones suficientes para poder obtener la imagen de la persona en la que se vean por completo sus brazos. Adicionalmente hay que garantizar que los colores del fondo no se confundan con los de usuario, por lo cual es conveniente usar una región de color homogéneo que contraste con el color de la piel y la ropa de la persona. No es necesario sacar de la imagen objetos que no se muevan, mientras estos contrasten con la persona o no queden entre la persona y el fondo desde el punto de vista de la cámara.

2.2 Substracción del Fondo

El primer paso fundamental es la segmentación del fondo y la persona. Los métodos recursivos y no recursivos [Pic04][Mci99][Che03], en los que se toman las últimas capturas para encontrar el fondo no son de mucha utilidad si la cámara permanece estática durante la actividad pues estos conllevan un alto costo computacional y son relativos al movimiento en la imagen. El método utilizado aprende una imagen de la escena, antes de que entre la persona, para diferenciarla con cada una capturas realizadas durante la actividad. El aprendizaje se logra realizando el promedio de 100 cuadros de la escena, lo que es conveniente de realizar por medio de un acumulador de *Running Averages* [Mci99] con lo que se requiere menor uso del buffer para obtener el promedio por ser un método recursivo. La diferencia, en valor absoluto, se compara contra un nivel Th de mínimo cambio, con el fin de darle un margen a los posibles cambios de luz y asimilación de sombras tenues.

$$|Captura - Fondo| > Th \quad (1)$$

Los píxeles que cumplan con esta condición se asumen como objeto y son marcados con 1, mientras que los que no la cumplan se asumen estáticos y son marcados con 0, obteniendo la silueta de la persona, quien debe ser lo único que se mueve en la escena. Este método es una ventaja pues permite acondicionar fácilmente cualquier espacio, con solo despejar el área de acción, más no todo el campo de visión de la cámara.

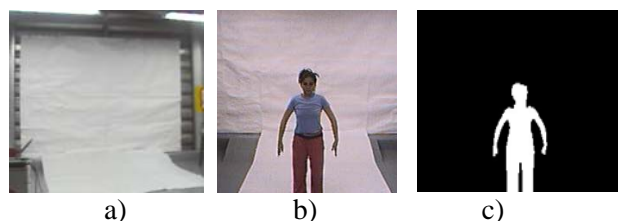


Figura 3. Substracción de fondo, a) fondo, b) actividad, c) silueta.

2.3 Esqueletización

Para realizar la esqueletización se asume que la persona está representada por los píxeles blancos y el fondo por los píxeles negros en la imagen de entrada. Se realizan adelgazamiento sucesivos [Par94] [Tsa05] [Tom02], encontrando los bordes según la evaluación de número de vecinos $B(p)$ y conectividad $X(p)$ de cada píxel para encontrar y borrar píxeles de la silueta en cada barrido. Seguidamente se implementa un par de condiciones, desarrolladas en este trabajo, que evitan borrar los píxeles que conforman el esqueleto. Para estas es necesario definir el vecindario de un píxel, numerando el vecino superior izquierdo como 1 y continuando en sentido de las manecillas del reloj. Se evalúa tanto en la silueta como en la imagen donde se marcan los candidatos de un ciclo.

P1	P2	P3
P8	Pi	P4
P7	P6	P5

C1	C2	C3
C8	Ci	C4
C7	C6	C5

a) b)

Figura 4. Vecindario para píxeles en a) la silueta, b) la mascara de candidatos.

El adelgazamiento y las condiciones que mantienen unido el esqueleto forman un total de 4 condiciones, las cuales son:

1. $2 < B(p) < 5$
2. $X(p)=1$
- Si $B(p)=5$ no borre cuando:
 3. $(P3 + P4 + P5) = 0 \ \&\& \ C8 = 0$
 4. $(P5 + P6 + P7) = 0 \ \&\& \ C2 = 0$

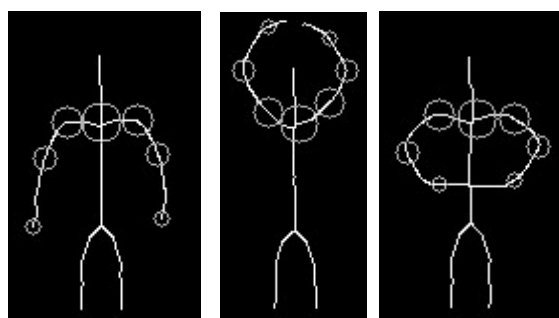


Figura 3. Esqueletos con el método propuesto.

2.2.1 Filtrado

Es necesario para obtener buenos resultados filtrar la silueta para eliminar píxeles ruidosos en el contorno y en el interior de la silueta, puesto que estos píxeles suelen generar falsas ramificaciones en el esqueleto. Un buen método encontrado es realizar un filtrado morfológico en el que se realiza una iteración de apertura (*opening*), para eliminar los píxeles del contorno, seguida de una iteración de cerramiento (*closing*), que cierre los huecos producidos por ruido dentro de la silueta.

2.4 Descripción de los brazos

A partir del esqueleto de la silueta de una persona es posible encontrar la ubicación de los puntos del pecho, los hombros, las manos y los codos que describen la fisonomía de las extremidades superiores del cuerpo humano. En el esqueleto, el pecho es un punto que en general tiene 4 ramificaciones, teniendo buenos resultados si se examina un perímetro de 7×7 alrededor de cada píxel. Los tramos de esqueleto que representan cabeza tronco y piernas no proporcionen información para la descripción de las extremidades superiores, por lo cual pueden ser borrados a partir de la ubicación del pecho trazando una línea, negra y vertical, del ancho del tronco. Este ancho se obtiene fácilmente del proceso de esqueletización, pues coincide con el número de ciclos requeridos para adelgazar esta parte del cuerpo, por ser la mas ancha en la imagen. Con esto se borra gran parte del ruido que pudiera aparecen en la región central del cuerpo. Así quedan solamente los tramos de esqueleto que representan los brazos. Con estos se determina la

ubicaron de los hombros y de las manos pues son los extremos de las líneas resultantes.

Para encontrar los codos se recorre la línea de los brazos de extremo a extremo, almacenando las coordenadas de cada píxel de la línea, para ubicar los codos por proporciones físicas del cuerpo humano aproximadamente a 2/5 de la distancia desde el hombro hasta la punta de los dedos. Con esto se hace indiferente encontrar el codo aunque los brazos estén extendidos o flexionados.



Figura 4. Brazos sin tronco con puntos marcados.

Seguidamente se calcula el ángulo formado entre los puntos del hombro y el codo y los puntos del codo y la mano, todos referidos al eje vertical de la imagen, tomando como cero grados la posición con los brazos relajados y aumentando a medida que los brazos suben hasta estar totalmente estirados arriba. El ángulo se toma igual para ambos brazos pero con signo cambiado, logrando magnitudes entre 0 y 180 grados para la articulación del hombro y entre -180 y 180 para la articulación del codo de cada brazo, ya que los antebrazos giran a partir del punto del codo, lo que permite cualquier dirección para estas partes de los brazos.

2.5 Reconocimiento de la postura

Para el reconocimiento se desarrolla un espacio de descriptores bidimensional, en el cual se cruzan los ángulos del hombro y el codo, asignando una región en el espacio para cada postura que se desee describir. Estas posturas son de los brazos ubicados de forma lateral y sólo realizando movimientos que no se desprenden del plano de la imagen. El análisis para ambos brazos es igual, teniendo en cuenta la simetría corporal, mas en el análisis es necesario cambiar el signo para los resultados de uno de los brazos. Este sistema toma como positivos los ángulos formados por el hombro izquierdo como se muestra en la tabla I.

TABLA I
CONDICIONES DEFINITIVAS PARA LOS ÁNGULOS MÁXIMO Y MÍNIMO DE CADA ARTICULACIÓN EN CADA POSTURA.

Ángulos Críticos	Hombro Min	Hombro Máx.	Codo Min	Codo Máx.
Postura 1 IZQ	0.0	40.0	-12.0	22.5
Postura 1 DER	0.0	-40.0	12.0	-22.5
Postura 2 IZQ	25.0	67.5	-22.5	-67.5
Postura 2 IZQ	-25.0	-67.5	22.5	67.5
Postura 3 IZQ	30.0	67.5	22.5	67.5
Postura 3 IZQ	-30.0	-67.5	-22.5	-67.5
Postura 4 IZQ	67.5	112.5	67.5	112.5
Postura 4 IZQ	-67.5	-112.5	-67.5	-112.5

Postura IZQ	5	67.5	130.0	157.5	-165.0 = 195.0
Postura IZQ	5	-67.5	-130.0	-157.5	165.0 = -195.0
Postura IZQ	6	112.5	157.5	112.5	157.5
Postura IZQ	6	-112.5	-157.5	-112.5	-157.5
Postura IZQ	7	145.0	180.0	157.5	-165.0 = 195.0
Postura IZQ	7	-145.0	-180.0	-157.5	165.0 = -195.0
Postura IZQ	8	112.5	170.0	-112.5	-170.0
Postura IZQ	8	-112.5	-170.0	112.5	170.0

III. RESULTADOS

Luego de las pruebas realizadas a más de 20 personas se puede concluir que con el sistema implementado se obtiene 100% de reconocimiento en todas las posturas, los problemas presentados se deben a que el usuario no hace la postura correctamente.

A partir de la medición de los ángulos para cada articulación es posible obtener graficas tanto en el espacio de descriptores como en tiempo. Con esto no solo es posible hacer un análisis de postura sino también un análisis dinámico del movimiento.

A partir de los resultados de los muestreos de todos los voluntarios se tomó el promedio de cada una de las posturas para todos y se realizó un promedio total, el cual se muestra en la figura 6 y se tabula en la tabla II.

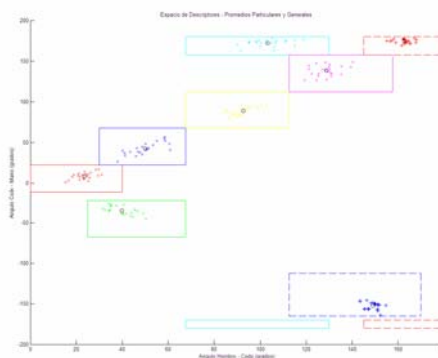


Figura 6. Descriptores promedio de los muestreos

TABLA II.
PROMEDIOS DE CADA POSTURA.

	Promedio Hombro	Ángulo Codo	Promedio Codo	Ángulo Hombro
Postura 1	23.5770	8.9084		
Postura 2	39.8778	-34.9328		
Postura 3	49.9286	42.4997		
Postura 4	92.5826	88.8341		
Postura 5	103.3152	172.1659		
Postura 6	128.6688	138.1117		
Postura 7	162.5891	174.4152		
Postura 8	149.0332	-152.6359		

Las curvas temporales se muestran en la figura 7.

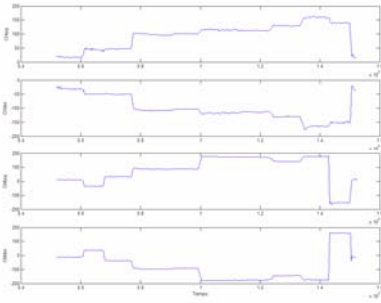


Figura 7. Graficas de ángulo contra tiempo del movimiento de las articulaciones de hombros izquierdo y derecho (chizq, chder) y codos izquierdo y derecho (cmizq, cmder).

Estas curvas fueron evaluadas y comparadas con un goniómetro biométrico electro-mecánico marca *Biometrix Ltd* [net-1], obteniendo resultados que nada tienen que envidiarle al método intrusivo.

La respuesta de sistema es de 125 ms en promedio, suficientemente rápida para ser considerada tiempo real en aplicaciones en las que no se necesita alta precisión o aplicaciones en las que el objeto analizado se mueve muy rápidamente. Pero, si se trata de entablar comunicación gestual con la máquina el tiempo de respuesta es más que suficiente.

Los promedios obtenidos para cada etapa se muestran en la tabla III

TABLA III
PROMEDIO TIEMPOS

	Imagen 640x480	Imagen 320x240
	Tiempo (ms)	Tiempo (ms)
Deteccion del Fondo	40	5
Filtrado Morfológico	<1	<1
Esqueletización	(Aprox 80 ciclos) 1000	120 (Aprox. 50 ciclos)
Busqueda del pecho	<1	<1
Busqueda de extremos	<1	<1
Busqueda codo	10	<1
Calculo angulos	<1	<1
Reconocimiento	<1	<1
Tiempo Total del Sistema	1050+(tiempo<5)	125+(tiempo<6)

El modelo del fondo, por ser recargable, permite tener más objetos en la escena, mientras estos no se muevan ni estén detrás del lugar en donde la persona realiza la actividad.

El algoritmo de esqueletización implementado, es más eficiente que el de Zhang & Suen, pues reduce la esqueletización en unas cuantas iteraciones. Por ejemplo una silueta esqueletizada en 73 ciclos por Zhang & Suen es esqueletizada en 68 ciclos por este nuevo algoritmo.

Usar dos grados de libertad para cada articulación es suficiente para hacer un análisis bidimensional de la disposición de los brazos de una persona.

El pecho es la única región sobre la imagen donde se encuentra el cruce de 4 ramas del esqueleto.

La aproximación de encontrar el codo midiendo 2/5 de la distancia entre el hombro y la punta de los dedos es satisfactorio y da buenos resultados.

El sistema es altamente invariante ante cambios en la escala y la ubicación lateral, mientras es poco invariante ante torciones del tronco.

IV. CONCLUSIONES

Se logró reconocer 8 posturas de las extremidades superiores de forma independiente para cada brazo, lo que implica un alfabeto total de 64 posturas del cuerpo donde todas están bien diferenciadas. Además es flexible en aumentar el número de posturas a reconocer, solo basta con incluir nuevas regiones en el espacio de descriptores.

Este sistema obtenido es apto para trabajar con cualquier persona, sin importar su edad, raza, sexo, o tamaño.

Con el método de descripción usando ángulos, además de lograr reconocer posturas, se obtiene un goniómetro útil para aplicaciones que buscan obtener un diagnóstico a partir del movimiento de las articulaciones superiores del cuerpo, en las áreas de fisioterapia o ergonomía, sin la necesidad de usar métodos intrusivos, los cuales son incómodos para la persona y que le impiden moverse libremente. Además puede ser implementado para el monitoreo de posturas que realizan las personas en su día a día, como lo son conductores, agricultores, secretarías y en general todo tipo de personal laboral.

Se obtuvo un algoritmo de esqueletización eficiente que proporciona esqueletos completos y sin falsas ramificaciones.

El sistema al ser invariante con respecto a la escala y a la traslación de la persona, permite que esta se desplace, por un espacio tan grande como el campo visual que la cámara lo permita y si esta decide usar solo una región del espacio, las áreas en la imagen no utilizadas en el movimiento pueden contener cualquier tipo de objeto, mientras este no se mueva ni cambie de color.

V. BIBLIOGRAFIA

[BRC99] Charles Rose, Bobby Bodenheimer and Michael F. Cohen. *Verbs and Adverbs: Multidimensional Motion Interpolation Using Radial Basis Functions*, 1999.

[CL04] Ana Grasielle Dionísio Corrêa, Roseli de Deus Lopes, *A Criatividade através da Expressão Musical: Uma Interface Gestual para Composição Musical Interativa*, 2004.

[NKS03] Katja Nummiaro, Esther Koller-Meier, Tomás Svoboda, Daniel Roth, and Luc Van Gool, *Color-Based Object Tracking in Multi-Camera Environments*, In Proceedings of the DAGM'03, Springer LNCS 2781, pp. 591-599, Sep 2003.

[Pic04] PICCARDI, Massimo, *Background Subtraction Techniques: a review*, University of Technology, Sidney 2004.

[Mci99] McIVOR, Alan M., *Background Subtraction Techniques*, Reveal Ltd, 1999.

[Che03] CHEUNG, Sen-Ching S. & KAMATH, Chandrika, *Robust techniques for background subtraction in urban traffic video*, Center for Applied Scientific Computing Lawrence Livermore National Laboratory, 2003.

[Tsa05] TSAI, Charlene , *Digital Image Processing Lecture 13: Image Topology – Skeletonization*, April 14, 2005.

[Par94] PARKER, Jim R., *Practical Computer Vision Using C*. Nueva York, 1994.

[Tom02] TOM, Kennant Kin Lun Performance evaluation of shape simplification methods, 2002.

[net-1] <http://www.biometricsltd.com/gonio.htm>