

MODELO ELIPSOIDAL EN 3D DE LA FORMA DE CAMINAR DE UNA PERSONA

Isidro Robledo Vega, Luis R. Salazar López, Jesús A. Domínguez Contreras

División de Estudios de Posgrado e Investigación

Instituto Tecnológico de Chihuahua

Chihuahua, México, 31310

{irobledo, lrsalazar, jadominguezc}@itchihuahua.edu.mx

RESUMEN

En este artículo se describe el diseño de un modelo en 3 dimensiones de la forma de caminar de personas con fines de reconocimiento. Para la generación del modelo tridimensional de una persona en movimiento es de suma importancia contar con datos confiables, para lo cual se desarrolló un sistema de visión en estéreo y los algoritmos necesarios para la reconstrucción de las escenas. El modelo se basa en la división de la persona en cinco regiones y el ajuste de elipsoides sobre los puntos localizados en cada una de ellas. Los rasgos descriptores son los centroides de los elipsoides y los ángulos de sus semiejes. Se realizaron pruebas de clasificación sobre una base de datos de 20 personas mostrando resultados prometedores.

1. INTRODUCCIÓN

El tema de la autenticación humana basada en métodos biométricos es un campo creciente dado el muy común fraude de suplantación de identidad [1]. Para reconocer a una persona se utilizan características fisiológicas (rostros, huellas dactilares e iris) y de comportamiento (forma de caminar, voz y firma escrita). Una de las ventajas de usar la forma de caminar como medida biométrica es que ofrece potencial para reconocimiento a distancia o a muy baja resolución o cuando otros métodos biométricos puedan no ser visibles [2]. Esto ayuda ya que en muchos casos es difícil distinguir el rostro de un individuo en una secuencia de video. Existen dos diferentes enfoques en el reconocimiento de la forma de andar: basados en modelos y libres de modelos. En el reconocimiento de la forma de caminar basada en modelos, la información como articulaciones o posición de la cadera se recopilan del cuerpo humano. Esta información es entonces utilizada para la construcción de un modelo para el reconocimiento [3][4]. En los métodos libres de modelos, las características de la forma de caminar son recopiladas de una secuencia de imágenes de siluetas binarizadas.

Utilizando estas siluetas, se puede recopilar la forma del caminar de las siluetas o reconstruir otra representación para el reconocimiento. Los enfoques libres de modelos en su mayoría adoptan las secuencias de las siluetas binarizadas, que son extraídas de un video utilizando técnicas de segmentación, tal como lo es la substracción de fondo. Sin embargo, la principal limitante para los enfoques libres de modelos es que requieren de imágenes de siluetas de muy buena calidad para trabajar.

En trabajos realizados previamente [5][6][7] se creó una base de datos de personas caminando, un conjunto de experimentos y un algoritmo base para la identificación de persona por su forma de caminar, el cual ha sido utilizado como referencia para probar múltiples técnicas desarrolladas por diferentes grupos de investigación. Aunque la base de datos contiene videos de personas caminando tomados desde dos cámaras no ha sido posible la reconstrucción tridimensional de las personas en movimiento debido a la geometría del sistema. En este trabajo estamos empezando una propuesta para la recolección de datos en 3D de las personas en movimiento y la generación de un modelo básico tridimensional de la forma de caminar de una persona. Nuestra hipótesis es que el uso de modelos en 3D, en lugar de las siluetas provenientes de una sola imagen, proporciona mayor información sobre las características del movimiento de cada persona y esperamos como resultado una mayor precisión al momento de la clasificación.

Iniciamos este trabajo con la adquisición de los datos utilizando un sistema de visión en estéreo, en otro trabajo de investigación [8] se diseñó una base de datos distribuida para el almacenamiento y administración de la información biométrica de personas en movimiento. Se realiza detección del movimiento sobre cada par de secuencias en estéreo y se reconstruye la persona en 3D utilizando solo los píxeles en movimiento, después se extrae un ciclo completo de la forma de caminar de la persona para

iniciar la construcción del modelo. Este consiste en dividir el cuerpo de la persona en cinco regiones y ajustar elipsoides sobre los puntos contenidos en cada una de ellas, utilizando como rasgos descriptivos las características de los elipsoides como son sus centroides y los ángulos de inclinación de los semiejes principales. Por último se realiza la clasificación de las personas basadas en estos rasgos. Cada una de estas etapas es descrita en las siguientes secciones.

2. SISTEMA DE CAPTURA

Para el desarrollo del sistema de captura de visión en estéreo se utilizaron los siguientes componentes: 2 cámaras marca Sony modelo Sony XCD-710CR cada una con un lente de longitud focal fija de 8.5mm, una base de aluminio para montar las cámaras y un trípode. Las cámaras tienen una interfaz Firewire (IEEE1394) y fueron conectadas a una estación de trabajo para captura y procesamiento de video digital. El área de captura dispuesta dentro de nuestro laboratorio se muestra en la Figura 1.

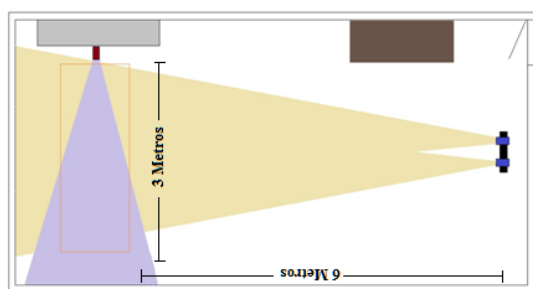


Figura 1.- Área de captura de video.

El sistema de visión en estéreo está encargado de la captura y análisis de las secuencias de imágenes adquiridas simultáneamente por las dos cámaras alineadas con una distancia de 10 centímetros entre ellas. A esta distancia se le llama Línea Base y es importante para los cálculos de reconstrucción tridimensional. Las personas caminan dentro del área marcada por el recuadro rojo en dos direcciones, de izquierda a derecha (LR) y de derecha a izquierda (RL). Ambas cámaras son sincronizadas utilizando un circuito de disparo por hardware. La adquisición sincronizada es controlada desde la estación de trabajo utilizando Matlab. En un principio se almacenaron los videos en la misma estación de trabajo pero al ir agregando sujetos a la base de datos nos obligó a desarrollar un esquema distribuido en servidores NAS para su almacenamiento y

administración [8]. Las cámaras utilizadas proporcionan el video en formato RAW, debido a la utilización de un filtro Bayer durante la captura. Los videos son convertidos por software en secuencias de imágenes RGB con formato PPM para su posterior procesamiento, un par de estas imágenes es mostrado en la Figura 2, en las líneas del piso se puede observar la diferencia en el punto de visión de ambas cámaras.



Figura 2.- Par de Imágenes en estéreo.

3. SEGMENTACIÓN DE MOVIMIENTO

La detección del movimiento humano y su seguimiento es el primer paso en el análisis de la forma de caminar de una persona. Lo primero que se asume es que las cámaras están estáticas y que el único objeto en movimiento en las secuencias es el sujeto caminando. El objetivo es determinar cuáles son los píxeles que se están moviendo entre cuadros de imagen consecutivos. Para esto se calcula el modelo del fondo de una secuencia de video, el método más simple es obtener el promedio de las n imágenes de la secuencia I . Este método es mejorado si se utiliza la mediana de cada píxel (x, y) de las imágenes en la secuencia para construir el modelo del fondo de forma que:

$$fondo_{xy} = mediana_t(I_{xy}^t)$$

Después se calcula la diferencia entre el modelo del fondo y cada una de las imágenes de la secuencia. La función para el cálculo de la diferencia es la siguiente [4]:

$$f(a, b) = 1 - \frac{2\sqrt{(a+1)(b+1)}}{(a+1) + (b+1)} \cdot \frac{2\sqrt{(256-a)(256-b)}}{(256-a) + (256-b)}$$

Donde a es el valor de un píxel en el cuadro de imagen y b es el valor del píxel en la imagen del fondo, ambos a y b corresponden a la misma posición (x, y) . Los valores de los píxeles pueden variar entre 0 y 255. Como estamos trabajando con imágenes RGB, este proceso se aplica a los tres canales por separado,

tanto en el cálculo del modelo del fondo como con la función de diferencia.

Se aplica un umbral para detectar si la diferencia en intensidad entre el pixel de una imagen y el modelo del fondo es suficiente para considerarlo parte del objeto en movimiento. Este umbral T puede ser calculado utilizando el método del histograma. Si el valor de la diferencia para un pixel (x, y) es mayor o igual al umbral en cualquiera de los tres canales (R,G, y B) entonces el pixel se considera parte del objeto, sino se considera parte del fondo. El resultado son imágenes de siluetas binarias obtenidas de acuerdo a:

$$Sil_{xy} = \begin{cases} 1 & f(a,b) \geq T \\ 0 & \text{De otra forma} \end{cases}$$

Se aplicaron operaciones morfológicas para mejorar la calidad de las siluetas binarizadas obtenidas en el proceso de segmentación de movimiento, en la Figura 3 se muestran las imágenes de ambas cámaras después del proceso de segmentación de movimiento.



Figura 3.- Par de siluetas binarizadas en estéreo.

4. RECONSTRUCCIÓN TRIDIMENSIONAL

El primer proceso para la reconstrucción tridimensional de una escena a partir de un par de imágenes en estéreo es el cálculo de la correspondencia entre puntos de la imagen izquierda con los de la imagen derecha. Si los planos de la imagen de ambas cámaras son coplanares, como ocurre con nuestro sistema de visión en estéreo, entonces la correspondencia se calcula solo sobre el mismo renglón de las dos imágenes. Se utilizó el algoritmo de emparejamiento global por bloques (SGBM) de las librerías de OpenCV, se hizo el ajuste de parámetros necesario para realizar el cálculo y se aplicó un filtro de suavizado sobre el mapa de disparidades resultante.

El siguiente paso es la identificación de los parámetros intrínsecos y extrínsecos del sistema de visión en estéreo. Debido a que se conoce la geometría del sistema se tiene conocimiento de todos

los parámetros por lo que se procede a al proceso de reconstrucción por triangulación descrito en [9] para lo cual usamos como referencia el marco de la cámara izquierda. Se utilizan las siluetas binarizadas como máscaras para reconstruir los puntos provenientes solo de los pixeles en movimiento en la escena que corresponden a la persona caminando. Los puntos en 3D obtenidos son guardados en un archivo PLY para su posterior procesamiento. La Figura 4 muestra una visualización de los puntos en 3D obtenidos en la reconstrucción por medio del software Meshlab.

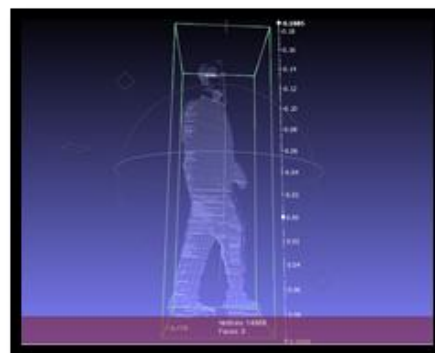


Figura 4.- Reconstrucción tridimensional de la persona caminando.

5. MODELADO DE LA FORMA DE CAMINAR EN 3D

Para iniciar la construcción de un modelo tridimensional de la forma de caminar primero se realizó la extracción de un ciclo completo de la forma de caminar de cada persona en la base de datos. Esta es una referencia importante ya que necesitamos diferenciar entre personas. Un ciclo completo de la forma de caminar se define como la secuencia única de las funciones por una extremidad que comienza cuando el pie de referencia (*ej.* pie izquierdo) hace contacto con el piso. La secuencia termina cuando se realiza un contacto subsecuente sobre el piso con el mismo pie. De esta forma, se realiza un proceso en el cual se detecta automáticamente el inicio y final del ciclo llevando a cabo un análisis sobre el ancho de las siluetas binarizadas, detectando el ancho máximo como inicio y dos máximos subsecuentes para el final.

El modelo se construye a partir de la división del cuerpo en cinco regiones: cabeza, brazo del lado derecho, brazo del lado izquierdo, pierna del lado derecho y pierna lado izquierdo. Se consideró el 20%

superior de los puntos para segmentar la cabeza, el 32% siguiente para los brazos y el 48% restante para las piernas, esto de acuerdo al eje Y (alto) y del 50% para separar las extremidades del lado derecho e izquierdo de acuerdo al eje X (ancho). Una vez hecha la división se ajustan elipsoides a los puntos en 3D en cada una de las regiones utilizando un algoritmo de envoltura convexa (*convex hull*). La Figura 5 muestra los elipsoides ajustados a las cinco regiones en que se divide a la persona.

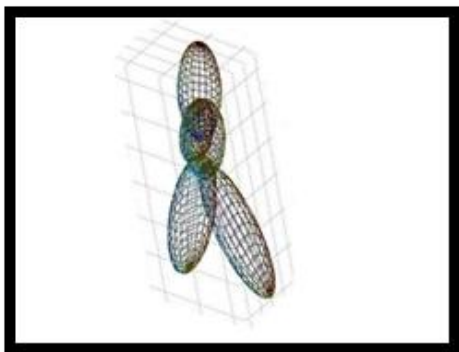


Figura 5.- Ajuste de elipsoides sobre las 5 regiones.

El vector de rasgos descriptivos se formó con los centroides normalizados y los ángulos de los ejes principales de cada elipsoide. De manera que se tiene una matriz de rasgos por persona, cada columna de la matriz es el vector de rasgos de un cuadro como se describió anteriormente. El número de columnas de la matriz depende del número de cuadros de la secuencia de un ciclo de la forma de caminar de cada persona. Las matrices de rasgos son proyectadas hacia un eigenspacio utilizando análisis de componentes principales (PCA), donde cada matriz es un punto en este eigenspacio, representando así el modelo de la forma de caminar de cada persona. Cabe mencionar que se utilizaron solo los eigenvectores asociados a los 20 eigenvalores más grandes para proyectar los datos hacia el eigenspacio de la forma de caminar ya que son los que representan la mayor parte de la información del modelo, logrando así reducir la dimensionalidad del problema de clasificación.

6. CLASIFICACIÓN DE PATRONES

Se cuenta con una base de datos de 20 personas y dos pares de secuencias en estéreo para cada una de ellas. Se utilizaron las secuencias en las que las personas caminan de izquierda a derecha (LR) para construir el

eigenspacio de la forma de caminar, a lo cual denominamos el proceso de entrenamiento. Después utilizamos las secuencias en la que las personas caminan de derecha a izquierda (RL) para el proceso de prueba. Como medida de similaridad se utilizó la distancia euclidiana entre los puntos de dos secuencias proyectadas al eigenspacio de la forma de caminar. La Figura 6 muestra la curva característica (ROC) con los resultados de la clasificación, donde podemos observar que se clasificó correctamente al 80% de las personas.

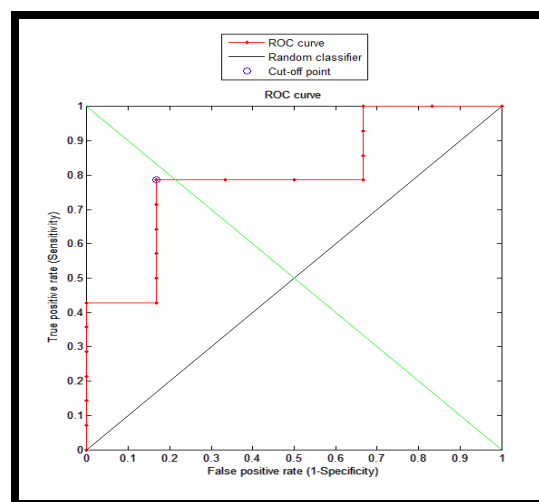


Figura 6.- Curva ROC con los resultados de la clasificación.

7. CONCLUSIONES

Se construyó un sistema de visión en estéreo para la captura de secuencias de video de personas caminando. Se formó una base de datos con videos de 20 personas caminando en dos direcciones. Se procesó la información en la base de datos para lograr la reconstrucción tridimensional de las personas caminando y se desarrolló un modelo para representar su forma de caminar con fines de reconocimiento.

Se desarrolló un esquema para la clasificación de las personas por su forma de caminar basado en el modelo desarrollado. Los resultados iniciales de la clasificación son prometedores. Actualmente nos encontramos revisando los rasgos considerados para la construcción del modelo de manera que podamos realizar ajustes que permitan mejorar los porcentajes de clasificación correcta. También realizaremos una comparación con clasificadores basados en modelos en dos dimensiones para determinar si el uso de

información tridimensional proporciona algún tipo de ventaja.

8. BIBLIOGRAFÍA

- [1] Prabhakar, S., S. Pankati and A. K. Jain, *Biometric Recognition: Security and Privacy concerns*. Proc. of the IEEE Security and Privacy, 2003.
- [2] Nixon, M., Tan, T., Chellappa, R., *Human Identification Based on Gait*. New York, NY: Springer Science, 2006.
- [3] Lee, L. & Grimson, W. E., *Gait Analysis for Recognition and Classification*, MIT Press, Cambridge, MA, 2002.
- [4] Wang, L., Tan, T., Ning, H., & Hu, W., *Silhouette Analysis-Based Gait Recognition for Human Identification*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(12), 1505-1518, 2003.
- [5] Sarkar, S., Phillips, P.J., Liu, Z., Robledo, I., Grother, P., Bowyer, K.W., *The HumanID Gait Challenge Problem: Data Sets, Performance, and Analysis*, IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 162-177, February, 2005.
- [6] Philips, P.J., Sarkar, S., Robledo, I., Grother, P., Bowyer, K., *The Gait Identification Problem: Data Sets and Baseline Algorithm*. International Conference on Pattern Recognition (ICPR 2002). Québec, Canada. August 2002.
- [7] Phillips, P.J., Sarkar, S., Robledo, I., Grother, P. and Bowyer, K., *Baseline Results for the Challenge Problem of Human ID Using Gait Analysis*, 5th IEEE International Conference on Automatic Face and Gesture Recognition (FGR 2002), Washington, D.C., USA, May 2002.
- [8] Cervantes, O.G., Robledo, I., Márquez, P.R., *Diseño de una Base de Datos Distribuida de Rostros y Forma de Caminar en 3D*, 33º Congreso Internacional de Ingeniería Electrónica, ELECTRO 2011, Instituto Tecnológico de Chihuahua, Octubre 2011.
- [9] Trucco, E., Verri, A., *Introductory Techniques for 3-D Computer Vision*, Ed. Prentice Hall, 1998.