

Detección de peatones con variaciones de forma al caminar con Modelos de Forma Activa

Pedestrian's detection with shape variations when walking with Active Shape Models

Juan Alberto Antonio

Universidad Autónoma del Estado de México, México

jaantoniov@uaemex.mx

 <http://orcid.org/0000-0003-3052-3171>

Marcelo Romero

Universidad Autónoma del Estado de México, México

mromero@uaemex.mx

 <http://orcid.org/0000-0002-4758-8484>

Recepción: 27 de marzo de 2019

Aprobación: 11 de febrero de 2020



RESUMEN

Se provee un detector de peatones con el algoritmo modelos de forma activa (ASM), con las etapas entrenamiento (PDM) y ajuste (ASM). Con PDM, se marcan 50 *landmarks* y se extraen los perfiles de grises en la silueta de cada peatón en 137 imágenes (peatón 1 y peatón 2) aplicando los modos de variación (PCA). El aporte de este trabajo es el ajuste y detección de un peatón a pesar de las variaciones. Al final con los resultados evaluados con *leave one out* en cada imagen de $1\ 080 \times 720$ píxeles y con la métrica del error cuadrático medio (MSE) se obtiene un promedio total de 12.7 píxeles en la distancia de error entre los *landmarks* originales y los *landmarks* estimados.

Palabras clave: modelo de formas activas, marcado, ajuste, variaciones de forma.

ABSTRACT

A pedestrian detector is provided with the algorithm models of active shape (ASM), with the stages: training (PDM) and adjustment (ASM). With PDM, 50 landmarks are marked, and gray profiles are extracted in the silhouette of each pedestrian in 137 images (pedestrian1 and pedestrian2) applying the variation modes (PCA). The contribution of this work is the adjustment and detection of a pedestrian despite the variations. At the end, the results evaluated with leave one out in each $1\ 080 \times 720$ pixels image and with the mean square error (MSE) metric, a total average of 12.7 pixels is obtained in the error distance between the original landmarks and the estimated landmarks.

Keywords: Active Shape Models, marking, adjustment, shape variations.

INTRODUCCIÓN

La detección de personas es de suma importancia para el estudio en muchas aplicaciones como la seguridad aeroportuaria, sistemas en centros comerciales, escuelas, entre otros (Fang *et al.*, 2017).

La detección de peatones (*PD, Pedestrian Detection*), como otra disciplina de la detección de personas, consiste en captar sujetos en diferentes poses de pie. Sus desafíos se basan en la gran variabilidad de la apariencia de los peatones debido a la pose, la vestimenta, la oclusión, condiciones de luz, el fondo y, en algunos casos, a las condiciones climáticas y de iluminación (Jordão y Schwartz, 2016; Angonese y Ferreira Rosa, 2017; Kim y Lee, 2013; Halidou *et al.*, 2014; Lakshmi, Faheema y Deodhare, 2016).

La detección de peatones utilizando cámaras de videovigilancia necesita de métodos que ayuden a capturar sus diferentes posturas al caminar, las cuales pueden ser cambiantes dependiendo de sus características (Dollár *et al.*, 2012) y por las variaciones de su cuerpo.

En la actualidad la detección de peatones se ha realizado a través de diferentes algoritmos, por ejemplo el uso de técnicas deformables como *Statics Shape Models (SSM)*, *Active Contour Models (ACM, Snakes)*, *Active Appearance Models (AAM)*, etc. (Hill, Thornham y Taylor, 2013; Blake, Curwen y Zisserman, 1993; Enzweiler y Gavrilu, 2009; Hilario *et al.*, 2005; Pentland y Horowitz, 1991; Vandenbroucke *et al.*, 1997; Zhang, Bauckhage y Cremers, 2015), los cuales se basan en el ajuste de las formas en las imágenes que contienen objetos por detectar. La problemática que representan estos métodos es la captura de la textura y además requiere una correspondencia y convergencia local mínima ante las variaciones de la forma (Flohr y Gavrilu, 2013).

En este artículo se propone el uso del Modelo de Formas Activas (ASM, *Active Shape Models*), el cual detectará y ajustará mediante *landmarks* (puntos característicos) escenas de peatones videograbados desde una cámara estática. De acuerdo con esto, la aportación que se pretende es el ajuste del ASM considerando la variabilidad al caminar, i. e., abriendo y cerrando los pies. Esta variabilidad es ajustada mediante una pirámide de resolución (Cootes *et al.*, 1995), la cual disminuye la cantidad de iteraciones necesarias para ajustar el modelo a pesar de las variaciones y los niveles de grises que se puedan encontrar, y que además es un tema poco explorado en el estado del arte (Das Choudhury y Tjahjadi, 2013; Müller y Arens, 2010; Ogawara, Li y Ikeuchi, 2007).

1. DETECCIÓN DE PEATONES CON EL MODELO DE FORMAS ACTIVAS

El estado del arte basado en ASM tiene una literatura amplia relacionada con el ajuste de formas en órganos del cuerpo como manos o rostros (Huysmans, Moens y Van Audekercke, 2005; Razali y Wahab, 2011; Le *et al.*, 2012). Sin embargo, en lo referente a la detección de peatones utilizando el algoritmo ASM, la literatura es limitada debido a que el uso de esta técnica es compleja y requiere un alto consumo de tiempo de cómputo (Cootes *et al.*, 1995); por este motivo, la gran mayoría de las investigaciones realizadas se basan en la detección de personas estáticas (pose de pie sin moverse), y en algunos trabajos se menciona cómo se detectan peatones mientras caminan, lo que permite analizar su cadencia (*gait recognition*).

Baumberg y Hogg (1994) demuestran cómo se puede dar seguimiento a un cuerpo no rígido en movimiento. Su detector funciona correctamente para poses y vistas aun con dos o tres peatones. La detección del peatón es mediante una silueta elipsoidal entrenada con 40 puntos característicos de la imagen (*landmarks*); no obstante, no muestran resultados de la efectividad de su propuesta.

Koschan *et al.* (2003) aplican ASM para detectar cuerpos no rígidos en una secuencia de video mediante una implementación jerárquica en espacios de color *RGB*, *YUV* y *HSI*. Marcan diferentes números de *landmarks* (10, 14, 21 y 42) y obtienen tres alineaciones de siluetas de una sola persona. Realizan detección de contorno y al final muestran el error normalizado entre los *landmarks* originales y los *landmarks* ajustados.

Jung (2008) propone un modelo con base en un exoesqueleto para poder detectar poses humanas. Usan sustracción de fondo y un algoritmo de coincidencia de ASM en el proceso de ajuste. Experimentan con 600 imágenes de personas con 17 *landmarks* que muestran características de pose. En los resultados obtienen tres componentes principales con un 80.62%, 93.3% y 97.32% de precisión.

Kim, Lee y Paik, (2009) identifican peatones en su forma de caminar a partir de una secuencia de 122 de personas obtenidas del conjunto de datos *HGCD* (Sadoghi Yazdi, Fariman y Roohi, 2012). Detectan humanos automáticamente con ASM utilizando 32 *landmarks* para la silueta de un humano a partir de imágenes de 720 x 480 píxeles. Al final obtienen una efectividad de 90% en la detección.

Lee y Choi (2011) detectan peatones utilizando ASM usando imágenes en infrarrojo (IR) y cámaras visibles que permiten seguirlos peatones en entornos degradados y con poca luz. Marcan manualmente 42 *landmarks* alrededor del cuerpo y al final no muestran resultados en sus experimentaciones.

Ma y Ren (2011) proponen un método basado en ASM para reconocer peatones. No especifican el número de *landmarks* para entrenar a sus peatones. Las imágenes con resolución de 320×240 píxeles fueron obtenidos de una cámara *RGB* omnidireccional. Al final muestran un 94% de éxito en la detección.

Ide (2013) combina ASM y *GrabCut* y lo llama segmentación de retroalimentación de forma (*ASFSeg*), el cual segmenta peatones. Este método compara resultados entre *GrabCut* y las muestras de ASM y elige la coincidencia para obtener la mejor segmentación. La tasa de error se muestra con *foreground (FG)* error con 2.32 % y *background (BG)* error con un 2.15 %.

Por su parte, Vasconcelos y Tavares (2015) detectan peatones que caminan en direcciones diversas. Para esto usan el conjunto de datos *CASIA Gait Database* (Arai y Andrie, 2012) obtenidos de videos en formato *MPEG* con escenas de peatones en imágenes de 320×240 píxeles. Entrenan 14 imágenes que representan la silueta del peatón y cada forma se representa por una serie de 113 *landmarks*, 100 *landmarks* de la silueta, combinado con 13 *landmarks* pertenecientes a codos, rodillas y pies. Sus resultados fueron de aproximadamente a un 95% de las imágenes de un peatón en movimiento en cuatro direcciones.

Por otro lado, en los trabajos relacionados con la detección de transeúntes no mencionan cómo ajustan la apertura completa del compás de las piernas (variación de forma al caminar), ya que muestran una segmentación de apertura de piernas poco representativa. En esta investigación se propone la detección de la variación al caminar con compás abierto o compás cerrado en una imagen en gris que contenga un peatón.

2. MODELOS DE FORMA ACTIVA

Los Modelos de Formas Activas (ASM) es un método flexible que ha sido usado para modelar y representar un amplio rango de objetos (Cootes y Taylor, 1992). En la primera etapa de ASM se realiza el modelo de distribución de puntos (PDM), el cual especifica la forma media del objeto modelado, además de las variaciones que pueda tener; para que el PDM sea robusto es necesario delimitar el contorno de la silueta mediante el marcado con los *landmarks* (figura 1). Adicionalmente, al modelo de distribución de puntos se añade uno de perfiles de grises basado en la obtención de los niveles de grises en cada *landmark* y así se podrá obtener en la segunda etapa del algoritmo ASM el ajuste a un nuevo objeto que contenga el perfil de gris más parecido a los niveles adquiridos en el entrenamiento.

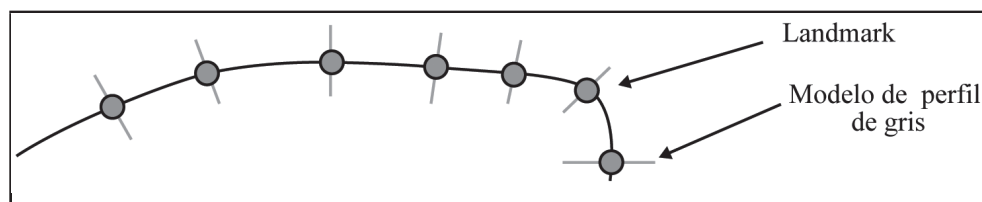


FIGURA 1

Representación del modelado de puntos (*landmarks*, color verde) y la obtención de los perfiles de grises (línea amarilla)

Fuente: Scott, Cootes y Taylor (2003).

A continuación, se muestran las etapas del algoritmo propuesto para la detección de peatones con variaciones al caminar, el cual se basa en el modelo de distribución de puntos de un conjunto de imágenes PDM y el ajuste a una nueva imagen ASM.

3. MODELO DE DISTRIBUCIÓN DE PUNTOS (PDM, POINTS DISTRIBUTION MODEL)

El modelo de distribución de puntos (PDM) consta de las siguientes etapas:

a) Marcado de *landmarks*: se obtienen con base en un marcado de puntos estratégicos alrededor de un objeto, el cual pertenece a un conjunto de datos de entrenamiento denotado por la ecuación:

$$x = [x_1, \dots, x_n, y_1, \dots, y_n]^T \quad (1)$$

b) Alineación del conjunto entrenamiento: se alinea el conjunto de entrenamiento mediante una matriz de transformación para buscar la mejor pose de los parámetros de forma y M la matriz de transformación.

$$X = M(s, \theta)[x] + x \quad (2)$$

donde:

$$M = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = s \begin{bmatrix} \cos\theta & \text{sen}\theta \\ -\text{sen}\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \quad (3)$$

$M(s, \theta)[x]$ alinea una rotación por θ , donde s es el parámetro de escalado y x el conjunto de entrenamiento de imágenes marcadas.

A partir de la ecuación 3 se obtiene la matriz M y posteriormente se genera el nuevo conjunto de entrenamiento alineado (X_i , ecuación 4)

$$X_i = Mx + (t_x, t_y) \quad (4)$$

c) Obtención de los componentes principales (PCA)

- Cálculo de la media de las m formas del conjunto de entrenamiento.

$$\bar{x} = \frac{1}{m} \sum_{i=1}^m X_i \quad (5)$$

- Cálculo de la matriz de covarianza S , del conjunto de entrenamiento.

$$S = \frac{1}{m} \sum_{i=1}^m (X_i - \bar{x})(X_i - \bar{x}) \quad (6)$$

- Construcción de la matriz donde $\phi_j, j = 1 \dots q$ son los eigenvectores de S .

$$\Phi = [\phi_1 | \phi_2 | \dots | \phi_q] \quad (7)$$

- Dado Φ y x , donde cada forma pueda aproximarse como:

$$x_i \approx \bar{x} + \Phi b_i \quad (8)$$

donde:

$$b_i = \Phi^T(x_i - \bar{x}) \tag{9}$$

d) Obtención de los perfiles de grises suavizados: el conjunto de niveles locales de grises (*LGL, Local Gray-Level*) son usados para capturar el nivel de gris local con su variación observado en cada *landmark* de la forma PDM (ecuación 10). A continuación se obtiene un modelo estadístico deformable donde $g_j, j = 1...n$ es el perfil de grises derivado y normalizado (gráfica 1).

$$f(g_{j,m}) = (g_{j,m} - \bar{g}_j)^T S_j^{-1} (g_{j,m} - \bar{g}_j) \tag{10}$$

3. 1. Ajuste a una nueva imagen: evaluación del modelo

3. 1. 1. Obtención de la pirámide de resolución

Cuando existen problemas con la longitud del perfil de gris, es necesario buscar el perfil más corto, donde la referencia del modelo debe estar cerca de su objetivo en la imagen antes del ajuste del modelo actual.

En caso de que los perfiles sean demasiado largos, la búsqueda se vuelve computacionalmente costosa y los niveles de gris se adhieren a otras estructuras alejadas del objeto de interés haciendo que ASM converja en la forma correcta. Debido a esto, se sugiere un enfoque de resolución múltiple de modo que la imagen tenga desde una baja hasta una alta resolución, generado con un suavizado gaussiano y submuestreo para producir una pirámide de resolución (figura 2). El nivel 0 de la pirámide es la imagen original y el nivel 1 es una imagen con la mitad del número de píxeles a lo largo de cada eje.

Después se utiliza una máscara gaussiana de 5×5 (se descompone linealmente en 2 circonvoluciones de 1-5-8-5-1) y luego submuestreando cada pixel del gris correspondiente.

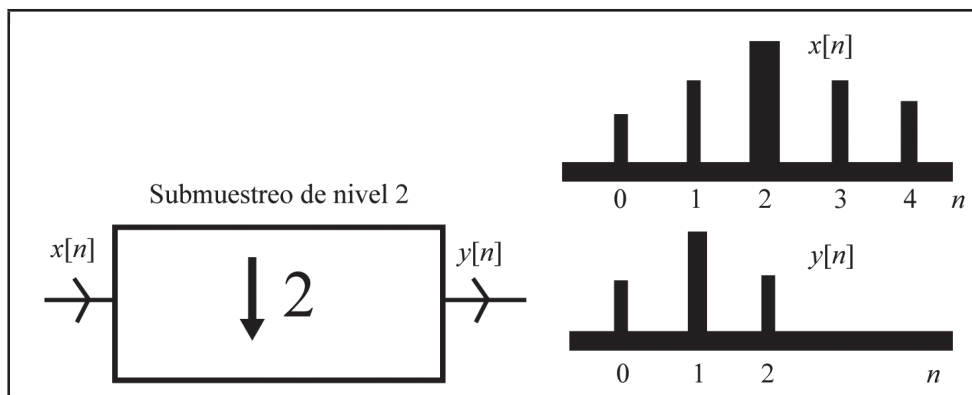


FIGURA 2

Ejemplo de submuestreo en el cambio de resolución en una pirámide de resolución de una imagen

Fuente: Scott, Cootes y Taylor (2003).

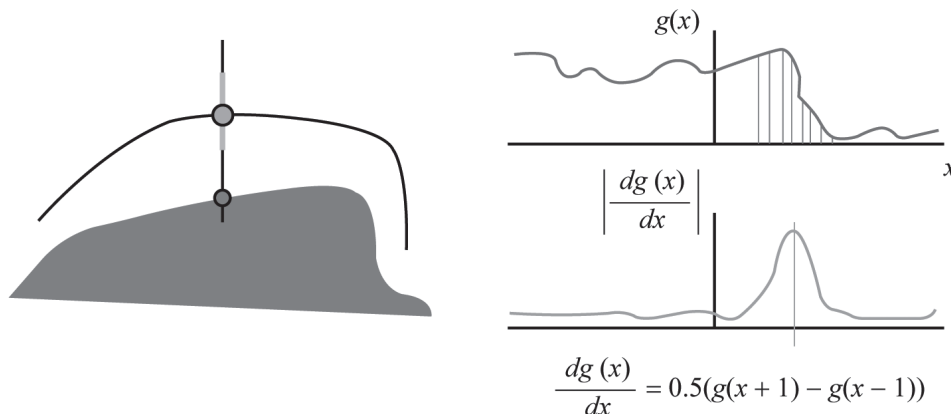
$$y[n] = \sum_n x[2n]e^{-i\omega n} \tag{11}$$

Posterior al ajuste de la nueva imagen con la pirámide resolución se obtiene el ajuste en base a los *landmarks* y los perfiles de grises:

$$b_g = P_g^T * y[n](g - \bar{g}) \tag{12}$$

$$g_{ajustado} = g + P_g b_g \quad (13)$$

$g_{ajustado}$ representa g desplazado por m muestras a lo largo de la dirección normal del límite correspondiente y con el cual se va midiendo la distancia para encontrar al perfil de gris representativo.



GRÁFICA 1

Búsqueda de los perfiles de grises en un nuevo modelo aplicando el perfil de gris que coincida con el nuevo objeto

Nota: Kim y Lee (2013)

De acuerdo con lo expuesto, un modelo estadístico deformable puede ser construido por la asignación de *landmarks* y la obtención de perfiles de grises. Para poder ajustar a una nueva imagen que no pertenece al conjunto de entrenamiento, se debe encontrar el grupo de parámetros de perfiles grises que mejor coinciden con el modelo de la imagen y que representa los límites del borde que corresponde a la silueta de un peatón.

4. PROCESO EXPERIMENTAL EN LA DETECCIÓN DE PEATONES CON VARIACIONES DE FORMA AL CAMINAR

En esta sección se describe de forma breve la propuesta para la detección de peatones con variaciones de forma al caminar, es decir, compás abierto y cerrado (figura 3). Básicamente, el algoritmo de la figura 3 consiste en dos fases: *a)* entrenamiento (PDM) y *b)* ajuste (ASM).

La fase PDM consiste en:

- a)* Alineamiento de las formas marcadas por los *landmarks* (ecuaciones 1-4).
- b)* Análisis de componentes principales (ecuaciones 5-9).
- c)* Construcción de niveles de grises (ecuación 10).

La fase de ajuste también conocida como ciclo ASM se puede describir como sigue:

- d)* Transformadas multirresolución aplicadas a las coordenadas de los contornos (ecuación 11).
- e)* Búsqueda activa (ecuaciones 12-13).

En general, el procedimiento de PDM y ASM propuesto en este trabajo se puede resumir con el Algoritmo 1. Para los fines de este artículo se ajustó el modelo de un peatón con 50 *landmarks*.

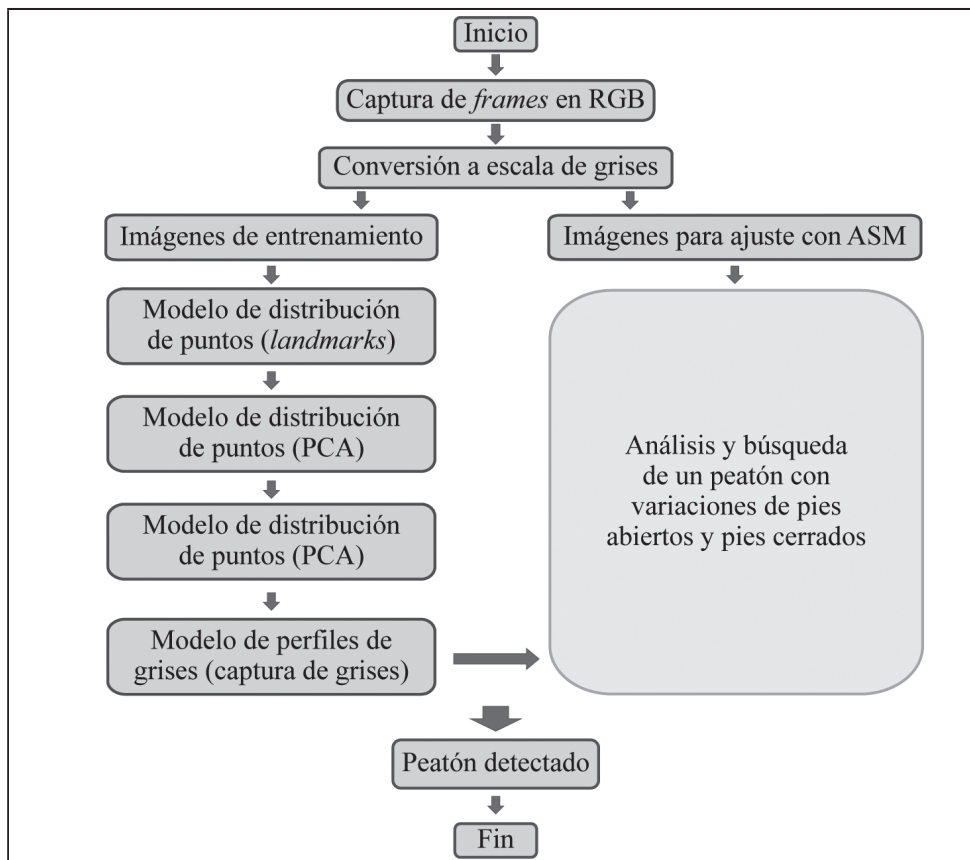


FIGURA 3

Etapas para la detección de un peatón con variaciones de forma con ASM

Fuente: elaboración propia.

Algoritmo 1 Algoritmo para la detección de peatones con variaciones al caminar
1: Seleccionar conjunto de imagenes de entrenamiento
2: landmarks = 50
3: for iteration = 1, 2, ... <i>Conjunto Entrenamiento</i> do
4: Marcar cada imagen con los landmarks predefinidos
5: Ec. (1)
6: end for
7: for iteration = 1, 2, ... <i>Total frames</i> do
8: Cálculo de Componentes Principales
9: {Ec. 2-9}
10: end for
11: for iteration = 1, 2, ... <i>Total frames</i> do
12: Algoritmo que busca la intensidad del modo del perfil de gris.
13: Pirámide de resolución ecuación (10)
14: Ec. (11)
15: for iteration = 1, 2, ... 40 do
16: Búsqueda y ajuste del perfil de gris adecuado
17: Ec. (12-13)
18: end for
19: end for

FIGURA 4

Algoritmo 1

Fuente: elaboración propia.

4. 1. Obtención de los datos experimentales

La obtención de los datos experimentales fueron a partir de la grabación de un video con escenas de peatones caminando de perfil para así poder obtener la apertura y cierre de piernas. El lugar donde se realizó la grabación fue en un pasillo de 4.50 m de largo \times 3.80 m de ancho en el interior del edificio de posgrado de la Facultad de Ingeniería de la Universidad Autónoma del Estado de México.

Se usó una cámara fija *RGB FaceTime HD* de 720p, la cual se localizaba a una distancia aproximada de 1.47 m de la zona donde caminaron los peatones. Para grabar a los peatones, se tomaron en cuenta varios factores tal como la cadencia de los pasos y la velocidad para poder obtener *frames* legibles y evitar imágenes borrosas que pudieran tener resultados de detección poco satisfactorios.

Todos los videos obtenidos se almacenaron en formato .mov con una resolución de 1 080 \times 720 pixeles y posteriormente se guardó en formato PNG. Los archivos de imagen PNG no pierden información al cambiarlos a escala de grises, lo cual es útil para ser usado por el algoritmo ASM. La tabla 1 muestra los conjuntos de entrenamiento necesarios en las escenas para el peatón 1 y peatón 2, donde la cantidad de *frames* se estableció en dos categorías para cada conjunto: *a*) pies abiertos y *b*) pies cerrados.

TABLA 1
Extracción de imágenes experimentales en peatón 1 y peatón 2

Cantidad de <i>frames</i>	Peatón 1		Peatón 2	
	Pies cerrados	Pies abiertos	Pies cerrados	Pies abiertos
<i>Frames</i> individuales	25	39	44	15
<i>Frames</i> totales	66		71	

Fuente: elaboración propia.

4. 2. Modelado de distribución de puntos

En la etapa de modelado de distribución de puntos se propuso un marcado de 50 *landmarks* alrededor del contorno del peatón (Godil, 2007) tomando en cuenta algunos puntos antropométricos según el modelo CAESAR (Ressler, 2001). De este modelo se pueden identificar los puntos específicos que corresponden a ciertas extremidades o partes del cuerpo humano, ya sea en modelos 2D o en 3D. En la figura 5, se muestran las imágenes de entrenamiento y su marcado (50 *landmarks*) correspondiente a los peatones con pies cerrados y con pies abiertos en las escenas de peatón 1 y peatón 2 para ser usados en las etapas de PDM (sección 4).

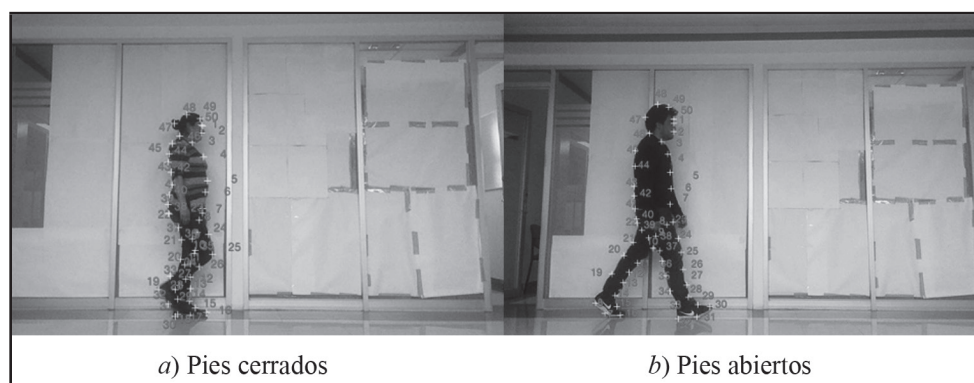
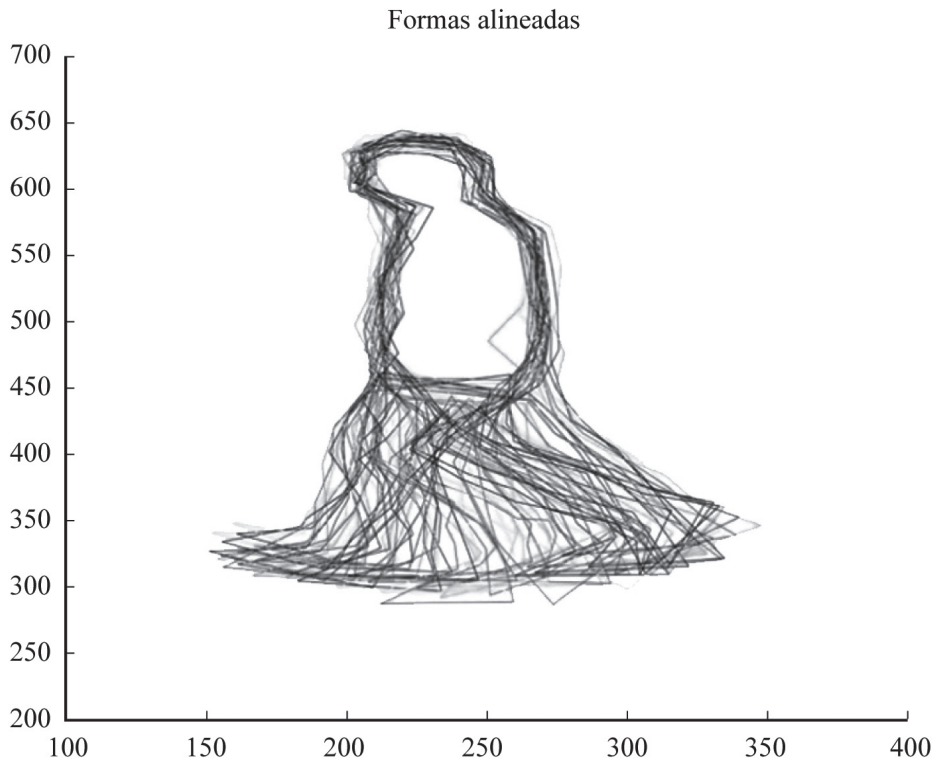


FIGURA 5

Etapas para la detección de un peatón con variaciones de forma con ASM

Fuente: elaboración propia.

Posteriormente, se busca la alineación del conjunto de entrenamiento (gráfica 2), y después se origina la forma media de todo el conjunto de entrenamiento (gráfica 3). Para la captura de perfiles de grises se utilizó una recta perpendicular a cada *landmark* de 40 píxeles de ancho (20 píxeles hacia arriba y 20 píxeles hacia abajo) y así utilizarlos en la búsqueda activa de grises en cada imagen del conjunto de entrenamiento. Se definieron como criterio de parada 40 iteraciones en la búsqueda del ajuste al mejor modelo de variación que se asemeje para poder segmentar la silueta del peatón tomando en cuenta las variaciones en el movimiento de las piernas al caminar.

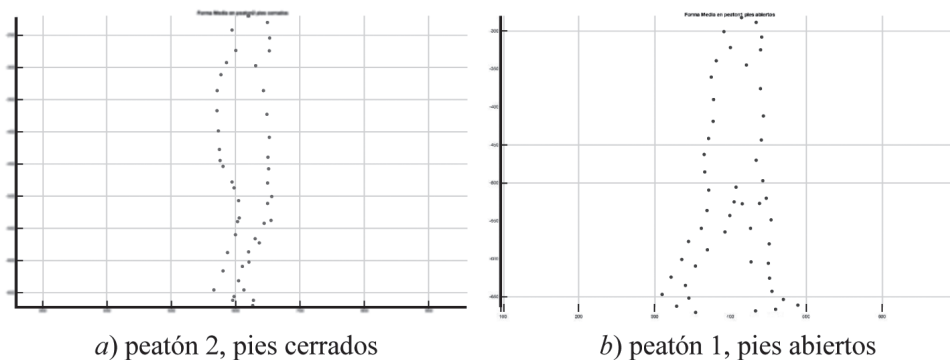


GRÁFICA 2

Representación de las formas alineadas en un peatón con variaciones de pies abiertos

Nota: elaboración propia.

La gráfica 3 muestra las formas medias del peatón 1 y peatón 2 con pies abiertos y pies cerrados con la técnica PCA.



GRÁFICA 3

Representación de las formas medias en a) y b)

Nota: elaboración propia.

5. EVALUACIÓN DE RESULTADOS

El algoritmo ASM fue desarrollado en MATLAB R2016b con un procesador a 2.3 GHz Intel Core i7 y una memoria RAM de 16 GB, y su ejecución ayudó a identificar y segmentar automáticamente el contorno de un peatón compuesto por 50 *landmarks* para posteriormente realizar el ajuste de la forma.

5.1. Métrica de evaluación

Como procedimiento de evaluación se utilizó la validación cruzada, en particular la técnica *leave one out* (Vehtari, Gelman y Gabry, 2017), la cual sirvió para estimar las imágenes (*test*) del conjunto de entrenamiento y dicha estimación fue medida con el método del error cuadrático medio (MSE). El procedimiento consistió en calcular la distancia entre los *landmarks* originales del marcado y los *landmarks* resultados del ajuste en la imagen de la prueba.

Los resultados de la búsqueda de ajuste con ASM en el peatón 1 (piernas abiertas) y peatón 2 (piernas cerradas) se muestran en la figura 6, en la cual se observa el marcado original (puntos amarillos) y el obtenido por el ajuste del ASM (puntos rojos). La mejor detección se realizó en el ajuste con piernas cerradas, donde la distancia de los puntos rojos está muy cercana al marcado original (puntos amarillos). En la imagen que muestra piernas abiertas (figura 6b) hay una leve separación de los puntos rojos con los puntos amarillos en los pies del peatón 1, lo que significa que en piernas abiertas el modelo ASM no se desempeña de manera óptima debido a la variación de la forma.

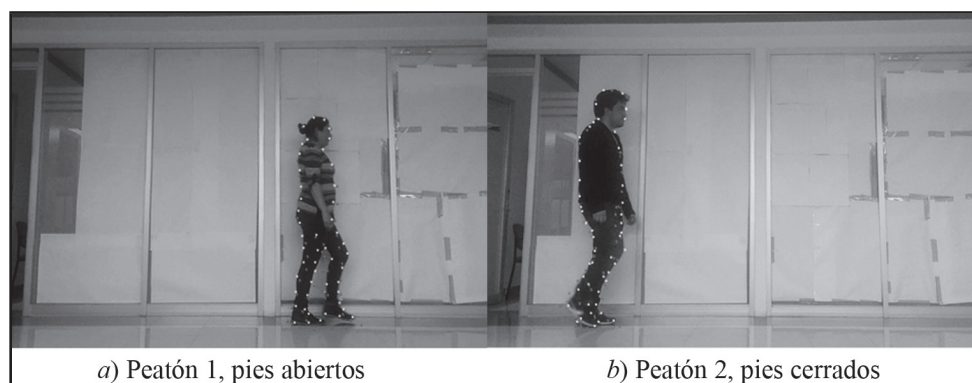


FIGURA 6

Representación del marcado original (puntos amarillos) y el marcado resultante del ajuste (puntos rojos)

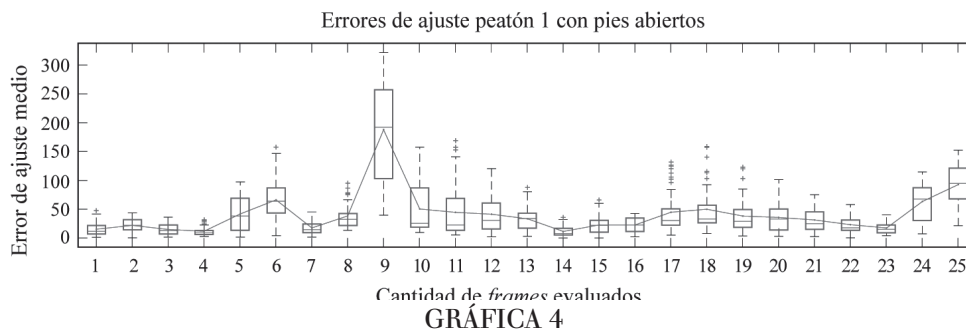
Fuente: elaboración propia.

5.2. Representación de los resultados obtenidos con ASM en los conjuntos de imágenes de peatón 1 y peatón 2

A continuación se muestran los resultados obtenidos (gráfica 4) con el procedimiento de validación cruzada y la técnica de *leave one out* cuando se usó como criterio de medida el MSE de la distancia entre el marcado original y el resultado de los *landmarks* ajustados en cada uno de los conjuntos de imágenes experimentales (peatón 1 y peatón 2 con piernas abiertas y piernas cerradas, respectivamente).

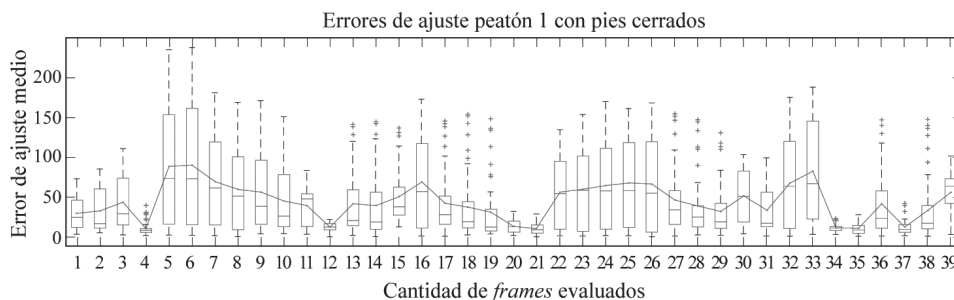
La gráfica 4 muestra el MSE del ajuste del ASM con los 25 *frames* que contienen la escena de peatón 1 (con piernas abiertas). Se observa que el peor resultado en el cálculo de la distancia con MSE entre los *landmarks* originales y los *landmarks* ajustados se obtiene con el *frame* 9 a una distancia de 150 píxeles, mientras que el ajuste más equilibrado se encuentra en el peatón 1, donde se observa que el *frame* 4 tuvo un resultado de

10 píxeles. El resultado promedio obtenido en la mayoría de los *frames* es representado por una línea roja donde se observa un MSE de distancia de 30 píxeles, por lo que se demuestra que el algoritmo ajustó las piernas abiertas en la mayoría de los *frames* a pesar de las variaciones por la apertura de las mismas.



Representación gráfica de diagrama de caja, donde se muestra la evaluación promedio del MSE en cada uno de los 25 *frames* del peatón 1 con pies cerrados
Nota: elaboración propia.

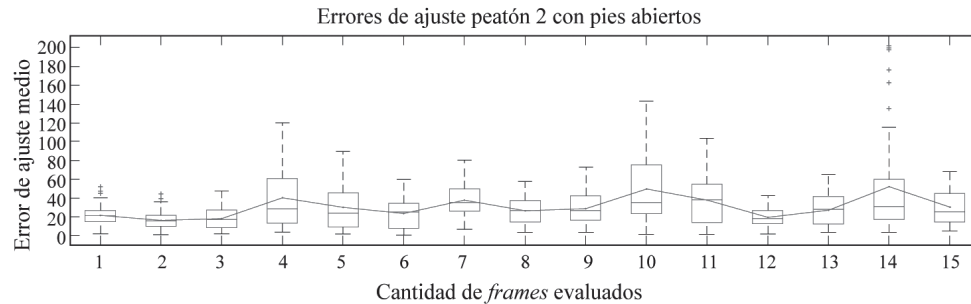
Por su parte, la gráfica 5 corresponde al conjunto de 39 imágenes experimentales del peatón 1 con pies cerrados y muestra mediante una línea roja que la distancia promedio de MSE es de 50 píxeles entre los *landmarks* originales y los *landmarks* resultado del ajuste. En algunos *frames* como 4, 12, 34, 35 y 37 se aprecia que la distancia de MSE es cercana a 0 píxeles demostrando en este caso que la detección del peatón 1 con pies cerrados contiene menos errores de ajuste que la detección con pies abiertos.



Representación gráfica de diagrama de caja, donde se muestra la evaluación promedio del MSE en cada uno de los 39 *frames* del peatón 1 con pies cerrados
Nota: elaboración propia.

La gráfica 6 corresponde a la evaluación de 15 *frames* del peatón 2 con pies abiertos. Muestra mediante la línea roja una distancia de MSE de 20 píxeles entre los *landmarks* originales contra los *landmarks* resultado del ajuste. En esta figura se evidencia la efectividad del método ASM, propuesto en este trabajo, en el peatón 2 con pies abiertos.

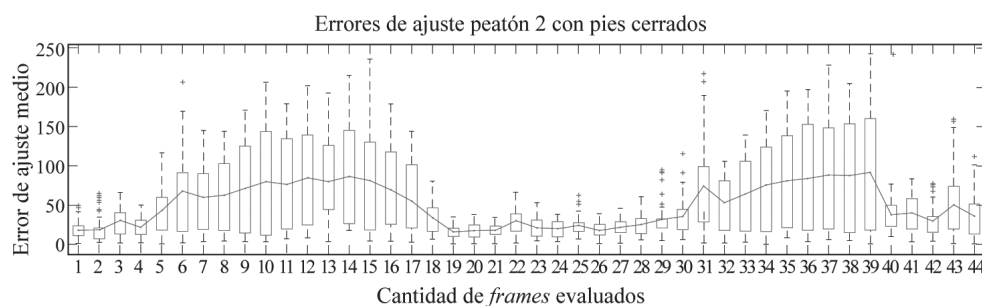
La gráfica 7 exhibe la evaluación de los 44 *frames* correspondientes al peatón 2 con pies cerrados. La línea roja señala el promedio de la distancia MSE, la cual es aproximada a 70 píxeles de separación entre los *landmarks* originales y los *landmarks* resultado del ajuste en 34 *frames*. En los *frames* 1, 4, 19, 20, 21, 23, 24, 25, 26, 27 y 28, del mismo conjunto, la distancia MSE es aproximadamente de 5 píxeles, siendo los *frames* con mejores ajustes. En otras palabras, en la mayoría de estos ajustes no se obtuvieron resultados satisfactorios; no obstante, en aquellos donde se tuvieron mejores resultados el ajuste fue bastante efectivo (distancia de MSE de 5 píxeles).



GRÁFICA 6

Representación gráfica de diagrama de caja, donde se muestra la evaluación promedio del MSE en cada uno de los 15 *frames* del peatón 2 con pies abiertos

Nota: elaboración propia.



GRÁFICA 7

Representación gráfica de diagrama de caja, donde se muestra la evaluación promedio del MSE en cada uno de los 44 *frames* del peatón 2 con pies cerrados

Nota: elaboración propia.

La tabla 2 muestra el promedio general del cálculo de la distancia MSE entre el marcado original y el marcado ajustado, los cuales fueron evaluados y mostrados en los diagramas de caja en las escenas de peatón 1 y peatón 2 con pies cerrados y pies abiertos. Se puede observar que el peatón 1 con pies cerrados y un conjunto de 39 *frames* tiene un promedio total de 41.6 píxeles del cálculo de la distancia MSE y con pies abiertos y un conjunto de 25 *frames* tuvo un promedio total de 12.7 píxeles en el cálculo de la distancia MSE. Por otra parte, el peatón 2 con pies abiertos y un conjunto de 15 *frames* tuvo un promedio total de 30.4 píxeles en el cálculo de la distancia MSE y con piernas cerradas y un conjunto de 44 *frames* tuvo un promedio total en el cálculo de la distancia MSE de 50.5 píxeles. Al finalizar estos resultados se puede decir que en algunos resultados la distancia resultado del ajuste (puntos rojos) parece estar muy separada de los *landmarks* originales resultado del marcado de entrenamiento (puntos amarillos), pero un factor a tomar en cuenta es la resolución de los *frames* evaluados, que en este caso fue de 1080×720 píxeles, con lo cual se puede concluir que ASM ajusta de manera correcta a pesar de las variaciones.

TABLA 2

Promedio total del error MSE de ajuste en peatón 1 y peatón 2 en cada conjunto de frames, ya sea con pies cerrados o con pies abiertos

	Peatón 1		Peatón 2	
Pose	Número de <i>frames</i>	Error medio de ajuste total	Número de <i>frames</i>	Error medio de ajuste total
Pies abiertos	25	41.6	15	30.4
Pies cerrados	39	12.7	44	50.5

CONCLUSIONES

En este artículo se propuso la detección de peatones con variaciones de forma al caminar usando el algoritmo ASM. Se evaluaron dos escenas denominadas peatón 1 y peatón 2, de las cuales se obtuvieron dos conjuntos de *frames* para cada escena (pies abiertos y pies cerrados). Se entrenó a cada conjunto de imágenes mediante el modelado de distribución de puntos con 50 *landmarks* alrededor del contorno de la silueta de cada peatón y en una etapa posterior se aplicó el ajuste de forma. Para poder evaluar el desempeño del ajuste se utilizó una validación cruzada, con la técnica *leave one out*, la cual sirvió para estimar las imágenes del conjunto de entrenamiento midiendo la distancia con el método del error cuadrático medio (MSE) entre los *landmarks* originales y los *landmarks* resultados del ajuste. Al final se muestran los mejores resultados dando un promedio total de 12.7 píxeles en la distancia MSE entre los *landmarks* del marcado original y los *landmarks* resultado del ajuste en el peatón 1 con pies cerrados, y un promedio total de 30.4 píxeles en la distancia MSE entre los *landmarks* del marcado original y los *landmarks* resultado del ajuste en el peatón 1 con pies abiertos.

ANÁLISIS PROSPECTIVO

Dado que se necesitan medidas de seguridad en los sistemas de videovigilancia, es necesario utilizar videos que capturen imágenes en vivo y que se puedan obtener de los centros de control, comando, comunicación, cómputo y calidad (C5) de cualquier entidad gubernamental de México. Con ello, se pueda entrenar y detectar peatones con variaciones al caminar con el algoritmo (ASM) y combinándolo con la fortaleza de otro algoritmo pueda disminuir los problemas de oclusión, iluminación y fondos difíciles. Se espera a futuro que los centros de control (C5) implanten el detector para ubicar peatones sospechosos, sobre todo porque los detectores de peatones modernos ubican y representan gráficamente la detección mediante un rectángulo (*bounding box*) y no mediante un modelo que ajuste la silueta.

AGRADECIMIENTOS

Se agradecen los comentarios de los árbitros de la revista que mejoraron sustancialmente el contenido del artículo.

REFERENCIAS

- Angonese, A. T., & Ferreira Rosa, P. F. (2017). Multiple people detection and identification system integrated with a dynamic simultaneous localization and mapping system for an autonomous mobile robotic platform. ICMT 2017-6th International Conference on Military Technologies, 779-786. <https://doi.org/10.1109/MILTECHS.2017.7988861>
- Arai, K., & Andrie, R. (2012). Gait recognition method based on wavelet transformation and its evaluation with Chinese Academy of Sciences (CASIA) gait database as a human gait recognition dataset. Proceedings of the 9th International Conference on Information Technology, ITNG 2012. <https://doi.org/10.1109/ITNG.2012.164>
- Baumberg, A. M., & Hogg, D. C. (1994). An efficient method for contour tracking using active shape models. Proceedings of 1994 IEEE Workshop on Motion of Non-rigid and Articulated Objects. <https://doi.org/10.1109/MNRAO.1994.346236>
- Blake, A., Curwen, R. y Zisserman, A. (1993). A framework for spatiotemporal control in the tracking of visual contours. *International Journal of Computer Vision*. <https://doi.org/10.1007/BF01469225>

- Cootes, T. F., & Taylor, C. J. (1992). Active Shape Models-‘Smart Snakes’. *BMVC92*. https://doi.org/10.1007/978-1-4471-3201-1_28
- Cootes, T. F., Taylor, C. J., Cooper, D. H., & Graham, J. (1995). Active shape models-their training and application. *Computer Vision and Image Understanding*. <https://doi.org/10.1006/cviu.1995.1004>
- Das Choudhury, S., & Tjahjadi, T. (2013). Gait recognition based on shape and motion analysis of silhouette contours. *Computer Vision and Image Understanding*. <https://doi.org/10.1016/j.cviu.2013.08.003>
- Dollár, P., Wojek, C., Schiele, B., & Perona, P. (2012). Pedestrian detection: An evaluation of the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. <https://doi.org/10.1109/TPAMI.2011.155>
- Enzweiler, M., & Gavrila, D. M. (2009). Monocular pedestrian detection: Survey and experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. <https://doi.org/10.1109/TPAMI.2008.260>
- Fang, F., Qian, K., Zhou, B., & Ma, X. (2017). Real-Time RGB-D based People Detection and Tracking for Mobile. *Proceedings of 2017 IEEE International Conference on Mechatronics and Automation*, 1937-1941.
- Flohr, F., & Gavrila, D. (2013). PedCut: An iterative framework for pedestrian segmentation combining shape models and multiple data cues. *Proceedings of the British Machine Vision Conference 2013*. <https://doi.org/10.5244/C.27.66>
- Godil, A. (2007). Advanced human body and head shape representation and analysis. *Digital Human Modeling*, 92-100.
- Halidou, A., You, X., Hamidine, M., Etoundi, R. A., Diakite, L. H., & Souleimanou. (2014). Fast pedestrian detection based on region of interest and multi-block local binary pattern descriptors. *Computers and Electrical Engineering*. <https://doi.org/10.1016/j.compeleceng.2014.10.003>
- Hilario, C., Collado, J. M., Armingol, J. M., & De La Escalera, A. (2005). Pedestrian detection for intelligent vehicles based on active contour models and stereo vision. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/11556985_70
- Hill, A., Thornham, A., & Taylor, C. J. (2013). *Model-Based Interpretation of 3D Medical Images*. <https://doi.org/10.5244/c.7.34>
- Huysmans, T., Moens, P., & Van Audekercke, R. (2005). An active shape model for the reconstruction of scoliotic deformities from back shape data. *Clinical Biomechanics*. <https://doi.org/10.1016/j.clinbiomech.2004.06.015>
- Ide, I. (2013). Segmentation of Human Instances Using Grab-cut and Active Shape Model Feedback. *Computer Science*, 11-14.
- Jordão, A. y Schwartz, W. R. (2016). *The Good, The Fast and The Better Pedestrian Detector*. Universidade Federal de Minas Gerais-Departamento de Ciência da Computação, 1, 1-51. Retrieved from <http://gibis.unifesp.br/sibgrapi16/eproceedings/wtd/19.pdf>
- Jung, C. J. (2008). *Human Pose Estimation ASM*. Retrieved from scholar.waset.org/1999.7/12456
- Kim, D., Lee, S., & Paik, J. (2009). Active shape model-based gait recognition using infrared images. *Communications in Computer and Information Science*, 61(4), 275-281. https://doi.org/10.1007/978-3-642-10546-3_33
- Kim, D. S., & Lee, K. H. (2013). Segment-based region of interest generation for pedestrian detection in far-infrared images. *Infrared Physics & Technology*. <https://doi.org/10.1016/j.infrared.2013.08.001>

- Koschan, A., Kang, S., Paik, J., Abidi, B., & Abidi, M. (2003). Color active shape models for tracking non-rigid objects. *Pattern Recognition Letters*. [https://doi.org/10.1016/S0167-8655\(02\)00330-6](https://doi.org/10.1016/S0167-8655(02)00330-6)
- Lakshmi, A., Faheema, A. G. J., & Deodhare, D. (2016). Pedestrian detection in thermal images: An automated scale based region extraction with curvelet space validation. *Infrared Physics & Technology*. <https://doi.org/10.1016/j.infrared.2016.03.012>
- Le, V., Brandt, J., Lin, Z., Bourdev, L., & Huang, T. S. (2012). Interactive facial feature localization. *Lecture Notes in Computer Science* (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-642-33712-3_49
- Lee, D., & Choi, S. (2011). *Multisensor fusion-Based object detection and tracking using Active Shape Model*. 2011 6th International Conference on Digital Information Management 2011, 108-114. <https://doi.org/10.1109/ICDIM.2011.6093321>
- Ma, J., & Ren, F. (2011). Detect and track the dynamic deformation human body with the active shape model modified by motion vectors. *2011 IEEE International Conference on Cloud Computing and Intelligence Systems*, 587-591. <https://doi.org/10.1109/CCIS.2011.6045137>
- Müller, J., & Arens, M. (2010). Human pose estimation with Implicit Shape Models. ARTEMIS'10-Proceedings of the 1st ACM Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Streams, Co-located with ACM Multimedia 2010. <https://doi.org/10.1145/1877868.1877873>
- Ogawara, K., Li, X., & Ikeuchi, K. (2007). Marker-less human motion estimation using articulated deformable model. Proceedings. *IEEE International Conference on Robotics and Automation*. <https://doi.org/10.1109/ROBOT.2007.363763>
- Pentland, A., & Horowitz, B. (1991). Recovery of Non-Rigid Motion and Structure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. <https://doi.org/10.1109/34.85661>
- Razali, N., & Wahab, A. (2011). 2D Affective Space Model (ASM) for detecting autistic children. *Proceedings of the International Symposium on Consumer Electronics*. <https://doi.org/10.1109/ISCE.2011.5973888>
- Ressler, S. (2001). A Web-based 3D Glossary for Anthropometric Landmarks. *Proceedings of HCI International*, 1, 1-5.
- Sadoghi Yazdi, H., Fariman, H. J., & Roohi, J. (2012). Gait recognition based on invariant leg classification using a neuro-fuzzy algorithm as the fusion method. *ISRN Artificial Intelligence*. <https://doi.org/10.5402/2012/289721>
- Scott, I. M., Cootes, T. F., & Taylor, C. J. (2003). Improving appearance model matching using local image structure. *Lecture Notes in Computer Science* (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-540-45087-0_22
- Vandenbroucke, N., Macaire, L., Vieren, C., & Postaire, J. G. (1997). Contribution of a color classification to soccer players tracking with snakes. *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*. <https://doi.org/10.1109/icsmc.1997.633237>
- Vasconcelos, M. J. M., & Tavares, J. M. R. S. (2015). Human motion segmentation using active shape models. *Lecture Notes in Computational Vision and Biomechanics*. https://doi.org/10.1007/978-3-319-15799-3_18
- Vehtari, A., Gelman, A., & Gabry, J. (2017). *Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC*. *Statistics and Computing*. <https://doi.org/10.1007/s11222-016-9696-4>
- Zhang, S., Bauckhage, C., & Cremers, A. B. (2015). Efficient pedestrian detection via rectangular features based on a statistical shape model. *IEEE Transactions on Intelligent Transportation Systems*. <https://doi.org/10.1109/TITS.2014.2341042>