UNIVERSIDAD AUTONÓMA DEL ESTADO DE MÉXICO

FACULTAD DE INGENIERÍA
DOCTORADO EN CIENCIAS DE LA INGENIERÍA

UNA METODOLOGÍA PARA DETECCIÓN DE MOVIMIENTO EN SECUENCIAS DE VIDEO: CASO DE ESTUDIO PEATONES

T E S I S

QUE PARA OBTENER EL GRADO DE:

DOCTOR EN CIENCIAS DE LA INGENIERÍA

PRESENTA:

M. en T.C. Juan Alberto Antonio Velázquez

DIRECTOR DE TESIS:
Dr. Marcelo Romero Huertas

Toluca, Estado de México, Noviembre, 2020

Resumen

La detección de peatones en secuencias de imágenes es un tema de investigación activo en el área de visión por computadora. Su estudio está motivado por el reconocimiento e interpretación automática de la detección del movimiento humano al caminar a partir de una escena de vídeo. Existe una variedad de aplicaciones donde la detección de peatones es importante, tales como: vigilancia para la seguridad inteligente, conteo de personas, análisis de movimiento al caminar, monitoreo e interpretación de vídeos deportivos, etc.
En los sistemas de videovigilancia, la detección de peatones tiene diversos factores que impiden que se realice una captura precisa. Éstas circunstancias pueden estar asociadas al medio ambiente y las características de la cámara, por ejemplo condiciones ambientales, resolución, campo de visión de la cámara y suspensión de la energía eléctrica. Por otro lado, la indumentaria de los peatones y la presencia de objetos móviles en la escena, son factores que reducen la efectividad de los sistemas de videovigilancia.
Por lo cual, este proyecto de investigación se basa en la implementación de una metodología basado en dos técnicas: detección de movimiento con sustracción de fondo (*DMSF*) y modelo de forma activa (*ASM*), el cual es aplicado en imágenes con peatones obtenidos desde una cámara de videovigilancia. El modelo de forma activa investigado en esta tesis, consta de 50 puntos de referencia alrededor de la silueta del peatón y una escala de grises de 40 *píxeles* en cada punto de referencia, el cual es utilizado para la detección de peatones en una escena. La principal contribución de este trabajo es la aplicación de la técnica de detección de movimiento y el ajuste de los peatones con *ASM*, que a pesar de las dificultades de fondo, escala, variaciones en la resolución y contraste entre el peatón y el fondo, da resultados prometedores.
El rendimiento del método propuesto en esta tesis se mide utilizando validación cruzada (*leave one out*), con dos conjunto de imágenes experimentales obtenidas de las base de datos *CDnet2014* y CASIA Gait dataset. Para medir el rendimiento del método propuesto se calcula el error de ajuste medio entre los puntos de referencia del *ground-truth* con los puntos de referencia estimados con el modelo de forma activa. Se obtuvieron los errores de ajuste al calcular la distancia euclidiana entre el *groundtruth* y los *landmarks* estimados por *ASM*. Al final se obtiene el mejor error medio de ajuste en el conjunto de datos CASIA Gait, con una puntuación de 4.5 píxeles

Tabla de contenido

Índice de figuras

Figura **Página**

Publicaciones

Parte de esta tesis ha sido publicada previamente en:

1. **Pedestrians' detection methods in video images: A literature review (2018)**. Juan A. Antonio y Marcelo Romero. On The International Conference on Computational Science and Computational Intelligence. Las Vegas NE, Estados Unidos de América, IEEE CPS.
   DOI 10.1109/CSCI.2018.00074.

2. **Detección de peatones con variaciones de forma al caminar con Modelos de Forma Activa (2020)**, Antonio, Juan Alberto; Romero, Marcelo. Aceptado para su publicación en la revista CIENCIA ergo-sum (ISSN: 2395-8782). Revista Científica Multidisciplinaria de Prospectiva de la Universidad Autónoma del Estado de México, la revista cuenta con una indización Open Journal Systems con los siguientes indíces: biblat, C.I.R.C EC3metrics, Clarivate Analytics, Clase, Dialnet, CONACYT, DOAJ, ERIHPLUS, ibss, iresie, MIAR, latindex, PKP INDEX, redalyc.org, REDIB, SHERPA RoMEO, SSOAR Social science y WorldCat.

3. **Pedestrian's localization into a video sequence using motion detection and active shape models (2020)**, J.A. Antonio, M. Romero and R.Alejo. Artículo enviado para su revisión en la revista IEEE Latin American Transactions (ISSN:1548-0992), la cual cuenta con una indización en Scopus, Science Citation Index Expanded, Aerospace Database, Civil Engineering Abstracts, Compendex, INSPEC, Metadex y Comunication Abstracts.

CAPÍTULO 1

Introducción

Los sistemas de vigilancia es un campo atractivo en los sistemas de visión por computadora y detección y seguimiento de personas [1]. La grabación de video se ha ocupado en los sistemas de videovigilancia (*CCTV*, por sus siglas en inglés) contando con un volumen considerable de imágenes que utilizan cámaras instaladas en ambientes urbanos, ahora es común verlos en diferentes lugares como parques, avenidas, terminales de autobuses y aeropuertos.
La videovigilancia basada en *CCTV* se ha desarrollado desde sistemas simples que comprenden una cámara conectada directamente a una pantalla de visualización, donde un vigilante se encuentra en una sala de control, observando incidentes de crimen o vandalismo o bien rastreando personas específicas, hasta complejos sistemas de múltiples cámaras conectadas a equipos de cómputo [2]. Estos sistemas necesitan operadores humanos que estén monitoreando eventos y que no retiren su atención de las pantallas, pero esto es difícil, ya que hay estudios que indican que un operador humano tiende a cansarse después de un determinado tiempo y con esta situación no se pueden detectar eventos que al parecer son importantes [3, 4].
Estos eventos o el análisis en línea es complicado debido a que los incidentes que existen en la vida real, tales como agresiones, raptos, golpes, manifestaciones entre otras no son captadas o analizadas por el observador.
La detección de peatones permite obtener información sobre las actividades que realizan los humanos a través del análisis de las características de sus trayectorias. El análisis de la posición y/o trayectoria de un individuo permite determinar si éste se encuentra caminando, corriendo, saltando, esperando algo, invadiendo una área no permitida, o bien desarrollando una actividad sospechosa. Es por eso que la detección de peatones es esencial para los sistemas de videovigilancia, ya que los transeúntes al caminar cambian de dirección, caminan a diferente velocidad con lo cual varía la forma del cuerpo con el movimiento de sus extremidades.
Debido a que las escenas donde caminan peatones existen problemas que hacen difícil la detección e identificación de una persona, tales como variaciones en el fondo de la escena, como movimiento, la iluminación y la indumentaria.

Para resolver parte de la problemática antes mencionada, se han propuesto técnicas como la detección de movimiento, la cual cuenta con diferentes alternativas para encontrar las regiones que cambian (*foreground*), diferenciándolo del fondo (*background*), la cual es efectiva a pesar de las dificultades que conlleva el desplazamiento.
En esta tesis se investiga el problema de la detección de peatones aplicado a secuencias reales de imágenes obtenidas desde una cámara fija. Para conseguir la detección se proponen dos técnicas alternativas, la primera es la técnica basada en detección de movimiento por sustracción de fondo (*DMSF*) el cual encuentra la región de movimiento, con lo que se obtienen los parámetros de traslación que sirve para activar una segunda técnica que es el modelo de forma activa del peatón.

1.1. Alcances

La investigación realizada en esta tesis se compara con el trabajo realizado por Vasconcelos et al. [5], el cual utiliza la base de datos *CASIA Gait dataset*, cuyas imágenes tienen una resolución de 340×240 *píxeles* y fueron capturadas a una tasa de 25 fps. Por el contrario, la investigación documentada en esta tesis, utiliza la base de datos *CDNet2014* que contiene imágenes con una resolución superior que contiene peatones caminando en diferentes direcciones. Una limitante del trabajo de Vasconcelos es el uso de imágenes con fondo (*background*) claro para generar un alto contraste, lo que facilita el ajuste del modelo de forma activa del peatón. Además, en los videos que contiene su base de datos experimental solo se percibe un peatón.
Esta tesis se centra, en el análisis de la detección de peatones, tomando en cuenta los problemas de fondo, para lo cual se seleccionó un conjunto de datos tomado de *CDnet2014*, el cual tiene variaciones de iluminación, posición y escala, que hace difícil la detección.
La aportación de esta tesis, es un método para detectar peatones en escenas de video, considerando la detección de movimiento y posteriormente el ajuste de modelos de forma activa de un peatón, el cual se ve robusto a pesar de la variabilidad al caminar, variaciones de fondo, bajo contraste, escala y la resolución de la imagen. Cabe mencionar que el ajuste del modelo de forma activa es automático, utilizando los parámetros de traslación calculados con la técnica de detección de movimiento.

1.2. Estructura de la tesis

Con base en los artículos 57, 59 y 60 del Reglamento de los Estudios Avanzados versión 2008, de la Universidad Autónoma del Estado de México, esta tesis de grado se desarrolla en la modalidad de tesis por artículo especializado y esta integrada por los siguientes capítulos:

Capítulo I Introducción, contextualiza el área de investigación estudiada en esta tesis.

Capítulo II Protocolo de investigación, presenta el protocolo de tesis registrado ante la Secretaría de Investigación y Estudios Avanzados.

Capítulo III Detección de peatones con variaciones de forma al caminar con Modelos de Forma Activa, muestra un artículo de investigación aceptado para su publicación en la revista CIENCIA ergo-sum.

Capítulo IV Pedestrian's localization into a video sequence using motion detection and active shape models, expone un artículo de investigación enviado para su revisión y posible publicación en la revista IEEE Latinoamerican Transactions.

Capítulo V Conclusiones y trabajo a futuro, concluye la investigación documentada en esta tesis y describe posibles lineas de investigación como trabajo futuro.

CAPÍTULO 2

Protocolo de investigación

2.1. Problemática

En esta tesis se plantea el problema de detectar peatones en un escenario con fondo complicado en escenas extraídas desde una cámara de videovigilancia. El problema consiste en estimar la ubicación de cada persona en cada *frame* de la secuencia y en determinar que el contorno encontrado corresponde a un peatón, desde que este entra hasta que éste sale de la escena, aún cuando se perturbe la apariencia del peatón durante la secuencia, con lo cual se obtengan falsos negativos y falsos positivos en la detección, y éstas pierdan su forma debido a variaciones.
Como se mencionó previamente la detección de eventos en línea son un campo atractivo en los sistemas de visión por computadora y la supervisión del tráfico [1], por esta razón existen sistemas de cámaras de vigilancia en diferentes lugares [6], pero la supervisión humana para la detección de eventos tiende a ser deficiente debido a que el humano se cansa después de treinta minutos [3, 4, 7]. Con base a lo anterior se necesita la automatización en la detección de peatones en escenas con movimiento es una necesidad y un reto para el desarrollo de la tecnología de videovigilancia [8, 9], debido a la complejidad de la ubicación de cada peatón en cada *frame* de la secuencia sin que se perturbe su apariencia.
En lo que se refiere al uso de las videocámaras (*CCTV*), en México, la instalación de 15,000 videocámaras de seguridad en la ciudad de México y 6,500 en el sistema de transporte colectivo metro [1], hacen que los sistemas *C5* puedan observar situaciones que ameriten dar seguimiento con las cámaras. En lo que se refiere a la seguridad a nivel nacional, los gobiernos estatales necesitaron instalar 25,631 videocámaras en toda la república mexicana en el 2015 [10],

[1] https://www.eleconomista.com.mx/politica/Se-necesitan-120000-camaras-para-cubrir-toda-la-CDMX-C5-20171110-0055.html

los cuales fueron instalados para supervisar la seguridad de los ciudadanos. Algunos problemas sociales que pueden atenderse con los sistemas de cámaras de videovigilancia son la detección de asaltos, agresiones, raptos, peleas, manifestaciones, etc [11, 12]. Sin embargo, para desarrollar dichas aplicaciones, la detección de peatones es una tarea esencial y preliminar a las etapas posteriores en aplicaciones específicas, por ejemplo, la detección el seguimiento de peatones [13, 14].
Más aún, la detección de transeúntes es complicado cuando se enfrenta a variaciones en iluminación, fondos, postura e indumentaria [15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26].
En esta tesis se investiga la técnica de detección de movimiento que servirá para localizar regiones candidatas que puedan contener peatones, a la cual se le puede ajustar un modelo de forma activa de un peatón. Logrando así la ubicación de cada peatón dentro de un *frame* de una secuencia de video.

2.2. Justificación

La detección de peatones en videosecuencias, es un problema estudiado por diversos investigadores [27], la cual es útil para el análisis de eventos tales como desplazamiento e interacción con el entorno. Actualmente, aparecen nuevas técnicas para detectar peatones donde el movimiento humano y el comportamiento han sido considerados [28].
En el estado del arte existen algoritmos para la detección de peatones, los cuales han tratado de resolver los problemas causados por diversas variaciones, por ejemplo el fondo, postura e indumentaria de los peatones [29].
Por lo cual, en esta tesis se investiga la localización de peatones en escenas de video capturadas por una cámara fija, donde dicha detección sea robusta ante las variaciones antes mencionadas.
Considerando que los sistemas de videovigilancia tienen su razón de ser para procurar la seguridad de la población. Por ejemplo, pensando en los peatones que transitan en la vía pública, se puede mejorar su seguridad analizando de manera automática las imágenes obtenidas por las cámaras. Procesando las imágenes capturadas por dichas cámaras, se podrían identificar diferentes eventos que contengan peatones. La relevancia de este análisis automático de las escenas es el poder identificar sucesos que puedan representar una amenaza a los peatones, por ejemplo, una aproximación sospechosa entre dos individuos que pueda consumarse como robo a un transeúnte, una pelea, o acoso. Luego entonces, la aplicación potencial de un análisis automático de las escenas de los sistemas de videovigilancia consiste en identificar dichas amenazas para alertar automáticamente a las instancias oficiales, a través de los centros de control, comando, comunicación, cómputo y calidad (C5). En resumen, con esta automatización no solo se evitaría que un operador esté observando los monitores 7×24 horas, reduciéndole la fatiga; además, se podrían detectar aquellos eventos que comprometan la seguridad de los peatones hoy en día.

2.3. Hipótesis

La detección de movimiento por sustracción de fondo, mejora la localización de peatones reportada en el estado del arte en secuencias de video.

2.4. Objetivo General

Implementar una metodología para localizar peatones mediante la detección de regiones de movimiento que permita ajustar modelos de forma activa.

2.4.1. Objetivos específicos

- Analizar las bases de datos del estado del arte para definir un conjunto de videos experimentales.
- Comparar algoritmos del estado del arte aplicables para detectar peatones en escenas de video.

2.5. Estado del arte

En este trabajo de investigación se estudia la detección de peatones basadas en dos técnicas de detección que son: detección de movimiento y detección por modelo de forma. En el presente capítulo se exponen los principales trabajos previos que proponen la detección de peatones con el uso de estas técnicas. Las ventajas y desventajas de estos enfoques así como las diferencias con este trabajo se describen a continuación.

2.5.1. Estado del arte en detección de movimiento

A continuación se presenta la literatura correspondiente a la detección de movimiento, donde se muestran algunas técnicas que son fundamentales para detectar regiones desde escenas de videovigilancia.
Sin embargo, la detección de movimiento sufre de diferentes problemas ocasionados por el ruido de origen, fondos complejos, variaciones en la iluminación de la escena y las sombras de los objetos estáticos y en movimiento. Por esta razón, se han sugerido varios métodos para superar estos problemas reteniendo solo el objeto móvil de interés. Éstos métodos han sido clasificados en tres categorías que son: sustracción de fondo, diferencia temporal y flujo óptico [30, 31, 32, 33, 34, 35, 36, 37, 38].
La diferencia temporal es altamente adaptativa y dinámica a ambientes; sin embargo muestra un bajo rendimiento al extraer todos los píxeles de características relevantes. Por lo tanto se aplican técnicas basadas en operaciones morfológicas y rellenado de huecos para obtener la forma deseada de un objeto en movimiento. Por otro lado la sustracción de fondo proporciona los datos de características más completos, pero es sensible a los cambios dinámicos debido a la iluminación y eventos extraños. Se ha categorizado la técnica de sustracción de fondo como: modelado básico de fondo, modelado estadístico de fondo, modelado de fondo difuso, agrupación de fondo, modelado de fondo por red neuronal, modelado de fondo *wavelet* y estimación de fondo [39, 40, 41, 42, 43, 44]. Por su parte el método de flujo óptico también puede ser usado para detectar objetos en movimiento. Sin embargo, la mayoría de los métodos de flujo óptico son computacionalmente complejos y no se pueden aplicar a transmisiones de video de *frames* completos en tiempo real sin tener un hardware especializado [45].
En [46], se muestra una técnica que mejora la diferencia de *frames*, al clasificar primero los bloques en el *frame* de fondo y después utilizando el coeficiente de correlación. Utilizan un método basado en clasificación de bloques a nivel de píxel para detectar el movimiento de personas en diferentes escenas de los *datasets* (*wallflower* e *I2R*), con un resultado promedio de precisión de 0.6396 %.
Por su parte en [47], presentan un enfoque robusto de inicialización de fondo basado en detección de movimiento de superpíxel. Las características de espacio y temporales de los *frames* se adaptan para eliminar los objetos en primer plano *foreground*. Primero seleccionan una subsecuencia con condición de iluminación estable para estimar el fondo, posteriormente segmenta las imágenes en superpíxeles para preservar la textura y los objetos de primer plano se eliminan mediante un filtrado. Utilizan el dataset *SBMnet*, el cual contiene 8 categorías del cual obtuvieron su mejor resultado en promedio *F-Measure* en la categoría *Low Framerate* con un resultado de 0.7222 a diferencia de otros métodos del estado del arte.
En [48], muestran un algoritmo basado en *W4* y diferencia de *frames*, que supera la deficiencia de detecciones falsas debido a mutaciones en el fondo; además de eliminar los huecos causados por la diferencia de *frames*. Este algoritmo es utilizado para detectar problemas de seguridad como la intrusión ilegal, alarma de persistencia y desplazamiento ilícito. Se analizaron 3 categorías infrarojas (*TM,SC y ROD*) tomadas de *DM642*, teniendo un resultado de 0.99 para la secuencia *DIN* y 0.9608 para la escena *ROA*.

En [49], proponen un método que busca una región de fondo dinámica analizando el video obtenido desde una cámara *CCTV* y que ayuda a remover falsos positivos. Éste fue evaluado con el dataset *CDnet 2012/2014*, teniendo un promedio de precisión en *CDnet2012* de 0.8650 y en *CDnet2014* un resultado de 0.7668.
En el trabajo realizado por [50], realizan un análisis de sustracción de fondo, diferencia de *frames* y el método *SOBS*; para detectar objetos en movimiento desde un vídeo obtenido de una cámara de videovigilancia *CCTV*. Al final no muestran resultados númericos, pero mencionan que el mejor método para encontrar movimiento es el de sustracción de fondo, ya que el método de diferencia de *frames* tiene un defecto el cual es no detectar a los objetos que tienen valores de intensidad distribuido uniformemente. Por su parte el método *SOBS* da buenos resultados, pero el tiempo de procesamiento es muy alto.
En una investigación basada en la detección de movimiento realizada por [51], quien presenta 12 técnicas basadas en detección de movimiento, en los cuales muestra los diferentes métodos de detección de movimiento como lo son: *temporal differencing (frame difference), three-frame difference (3FD), adaptive background (average filter), forgetting morphological temporal gradient (FMTG), Background estimation, spatio-temporal markov field, running gaussian average (RGA), mixture of gaussians (MoG), spatio-temporal entropy image (STEI), difference-based spatio-temporal entropy image (DSTEI), eigen-background (Eig-Bg) y simplified self-organized map (Simp-SOBS)*. Al finalizar el mejor resultado analizado con el *dataset CDnet2014*, es el evaluado por *GMM* con 0.99593 de especificidad, 3.08499 en *PWC* y 0.61021 de precisión. Mientras que la técnica de detección de movimiento con el menor rendimiento es *STEI*, con 0.78646 de especificidad, 22.18321 en *PWC* y 0.12881 de precisión.

2.5.2. Estado del arte en detección de peatones con modelado de forma

Los modelos de formas activas (ASM, por sus siglas in inglés), han sido utilizados en diferentes áreas de la visión computacional, para propósito de segmentación de imágenes, en áreas tales como la medicina, industria y seguridad. El problema de rastrear a las personas y reconocer sus acciones en secuencias de video es de creciente importancia para muchas aplicaciones [52, 53, 54]. Los ejemplos incluyen videovigilancia, interacción humano-computadora y captura de movimiento para animación, por nombrar algunos. Se requieren consideraciones especiales para el procesamiento de imágenes digitales al rastrear objetos cuyas formas (y/o sus siluetas), cambian entre *frames* consecutivos. Por ejemplo, los ciclistas en una escena de la carretera y las personas en la terminal de un aeropuerto pertenecen a esta clase de objetos denominados objetos no rígidos. Los modelos de forma activa (ASM) se pueden aplicar a la segmentación de objetos no rígidos en una secuencia de video. A continuación se describe la literatura relacionada al modelo de formas activas aplicada a la segmentación del contorno de un peatón.

Baumberg and Hogg [55], utilizan el *modelo de forma activa* para dar seguimiento a un cuerpo no-rígido en movimiento. El filtro de *Kalman* fue usado para controlar la escala espacial para la búsqueda de características en *frames* sucesivos. El rastreador de personas funciona correctamente para poses y vistas aún con 2 o 3 peatones incluso parcialmente ocluídos. Su rastreador solo funciona correctamente para poses y vistas que estén representadas en el conjunto de entrenamiento, además de no reconocer formas grandes. Al final muestran que (signal-to-noise ratio, SNR) es 0.05 con 26 dB.

En los sistemas en reconocimiento de pasos (*gait recognition*) Kim et al. [56], utilizan imágenes de infrarojos que ofrecen un rendimiento de reconocimiento de personas constantes aún a pesar de problemas de oclusión parcial o iluminación dinámica. Usan *modelos de forma activa* para obtener un reconocimiento robusto con fondo complicado, iluminación pobre y oclusiones y para esto crean un modelo de 30 puntos. Almacenan sus características en un vector 4D para posteriormente obtener las distancias en cada paso que da una persona. Un problema mencionado es que funciona con imágenes sin problema de fondo. Los resultados mencionan una relación de coincidencia *(matching-ratio)* de 93.8 % en imágenes obtenidas de una cámara infraroja y un 62 % en cámaras *CCD*.

Kim and Paik [57], presentan un estudio de *gait recognition* a partir de una secuencia de 122 siluetas de personas obtenidas de un video de baja resolución Sarkar et al. [58]. Reconocen humanos automáticamente a partir de la extracción de características basándose en los *modelos de forma activa*, además el algoritmo propuesto supera los inconvenientes de ruido, eliminación de sombras y una mayor tasa de reconocimiento. Para modelar cada peatón y

obtener sus puntos se basaron en la detección de personas mediante el *dataset HGCD* Sadoghi Yazdi et al. [59], utilizan 32 *landmarks* del cuerpo humano para poder detectar individualmente a cada peatón en una escena obtenida de un frame de video de 720×480 píxeles, una debilidad es no poder detectar al peatón caminando. Al finalizar hacen una comparación con el algoritmo *Base Line* [60], y mencionan una tasa de rendimiento de 97 % en vista, un 95 % en zapato, un 91 % en vista+zapato, 92 % en superficie, un 86 % en zapato+superficie, un 85 % en vista+superficie y un 78 % en vista+zapato+superficie.

Lee and Choi [61], proponen un sistema automático de detección y seguimiento de objetos utilizando el algoritmo *modelos de forma activa*. Utilizan sensores de infrarojo *(IR)* y cámaras visibles que permiten el seguimiento en entornos degradados y con poca luz. Cada conjunto de entrenamiento es localizado mediante una región elipsoidal con ayuda del algoritmo *SSA* Moskvina and Zhigljavsky [62]. El sistema puede rastrear una sola persona a partir de un marcado manual de 42 *landmarks* que ayudan a modelar el cuerpo en diversos entornos en tiempo real aún con oclusión. Usan *frames* de 320×240 píxeles extraídos de video-grabaciones en exteriores e interiores con una cámara de circuito cerrado y con problemas de iluminación. Una debilidad que muestra es si no haya parámetros de seguimiento para ASM, éste no puede llegar a ajustar a una persona, y otra dificultad es la pobre detección en fondos complicados. Al final no muestran resultados.

En Jang and Jung [63], estiman la pose humana con ayuda de los *modelos de forma activa* para la detección de la silueta de un peatón a partir de 600 imágenes. Las desventajas que se tienen en base a la precisión de la postura del ser humano cuando hay variación en brazos y piernas y cuando la silueta del cuerpo humano se presenta en forma desproporcionada. Usan un modelo en base a exo-esqueleto, el cual ayuda a detectar varias poses humanas entrenadas con 17 *(landmarks)*. Las problemáticas a las que se enfrentaron, fueron la detección en fondos difíciles y la piel y ropa de las personas. Al final obtienen una pose ideal donde se necesitaron 30 iteraciones en un tiempo inferior a 0.03 segundos.

En Koschan et al. [64], utilizan los *modelos de forma activa* para el seguimiento de un cuerpo no rígido en una secuencia de video con características de imágenes que no incluyen información de color e implementan un modelo jerárquico mejorado para el seguimiento de personas en una secuencia de video-vigilancia a color en espacios de color *RGB, YUV* y *HSI*. Entrenan peatones con (10, 14, 21 y 42 *landmarks*), elegidos manualmente en el contorno inicial y se obtienen 3 alineaciones de siluetas de una sola persona. Su algoritmo muestra algunas dificultades al enfrentarse a problemas de iluminación y al color, además de que solamente puede detectar una sola persona. Al finalizar tienen un mejor desempeño en la detección individual de peatones con 42 puntos de referencia y además de detectar aún con oclusión parcial con objetos.

En Arjun [65], proponen un algoritmo de estimación de pose humana para detectar una persona a partir de una imagen en 2D. Definen una postura humana a partir de una combinación de puntos que son significativos y que son invariantes bajo algunas transformaciones, específicamente seleccionan 14 *(landmarks)* en puntos anatómicos que incluyen articulaciones, como el codo y las rodillas; los puntos finales, como la cabeza, las manos y los pies. Estos puntos son utilizados para estimar un cuerpo en 3D. Después de definir una sola postura se localiza la región de la pose de la imagen del modelo tridimensional aplicando la correspondencia de distancia de *Hausdorff*. Al final muestran un resultado de 45 segundos en la ejecución del método propuesto.

En Vasconcelos and Tavares [5], detectan el movimiento de peatones utilizando el dataset *(CASIA gait database)* Zheng et al. [66], en formato de video *MPEG* con una resolución de 320×240 píxeles. Utilizan *point distribution model (PDM)* para un conjunto de entrenamiento de 2734 imágenes. Tienen 14 sujetos diferentes que caminaron en cuatro direcciones (0^o, 36^o, 54^o y 90^o) representada por 113 *(landmarks)*, conformados por 100 *landmarks* de contorno automáticamente extraídos de la silueta, además de 13 *landmarks* en codos, rodillas y pies. El algoritmo *PDM* fue utilizado posteriormente para construir un *modelo de forma activa*, combinado con el modelo de perfiles de niveles de grises, que ayudaron a ajustar el contorno del cuerpo humano. Los resultados fueron de 95 % de rendimiento y mencionan que su sistema funciona bajo condiciones controladas y bajo ambientes restringidos.

En Ma and Ren [67], realizan una modificación a los *modelos de forma activa (ASM)* para reconocer vectores en movimiento útiles para detectar un humano. El modelo de forma activa identifica las formas humanas desde un conjunto de entrenamiento, cabe mencionar que no especifican el número y posición de los *landmarks*. Entrenaron diferentes

tipos de modelos de cuerpos humanos modificados por vectores de movimiento obtenidos de secuencias de video *RGB* de 24 bits y con una resolución de 320×240 píxeles, capturados de una cámara omnidireccional de 360º VS-C15N. Utilizan un método de flujo óptico para obtener los vectores de movimiento. El algoritmo para ajustar debe utilizar el método de flujo óptico para obtener los vectores de movimiento y depende de la posición y postura del peatón a encontrar para ver su efectividad. Al finalizar muestran una tasa de reconocimiento de humanos de aproximadamente 94 %.
Por otro lado Pourjam et al. [68], mencionan un método que combina *active shape model* y *grab cut* y le llaman (*ASFSeg, shape feedback segmentation method*) el cual segmenta automáticamente humanos (peatones). Compara la máscara de resultados de segmentación de la etapa de *grab-cut* y las muestras generadas por *ASM* y luego elige la coincidencia más cercana para generar nuevas segmentaciones hasta converger con las máscaras generadas. Entrenaron 13 ejemplos de humanos para entrenar el modelo *ASM* y para poder segmentar aleatoriamente estos humanos. El principal problema que mencionan son las características de fondo y el color, además de tener un mayor número de entrenamiento para ajustar un nuevo modelo. La tasa de error es cercana al 1.5 % en imágenes de personas con mochilas y otros objetos como sombreros.

2.6. Metodología

Para el desarrollo de esta investigación se propone una metodología de tres etapas: Detección de movimiento, ajuste del modelo de forma activa y evaluación, la cual se muestra en la Figura 2.1. La primera etapa ubica la región de movimiento, donde se encuentra el objeto deseado donde se pudiera ubicar un peatón. La segunda etapa consiste en aplicar la técnica basada en modelos de forma activa, que ajusta un modelo humano a partir de 50 *landmarks*. En la tercera etapa se evalúa el rendimiento del método de detección de peatones.

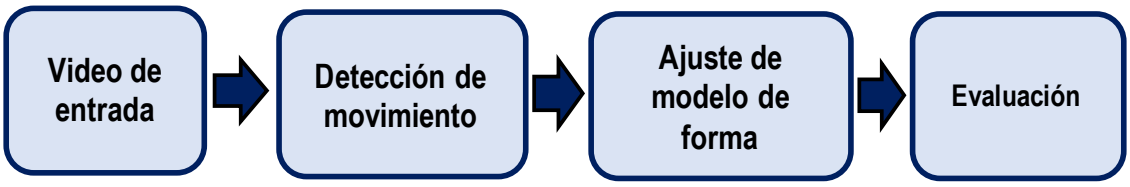


Figura 2.1: Etapas propuestas para la detección de peatones.

2.6.1. Detección de movimiento por sustracción de fondo (*DMSF*)

El método de detección de movimiento con sustracción de fondo (*DMSF*) mostrado en la Figura 2.2, detecta regiones de cambio entre dos *frames*, cuyas etapas son las siguientes:

1. Convierte el primer frame de la videosecuencia a escala de grises, y es definido como frame de fondo (F).
2. Lee el siguiente frame en la videosecuencia (I) y se convierte a escala de grises.
3. Sustrae el fondo F de la imagen I, $B(x, y) = I(x, y) - F(x, y)$.
4. Binariza B, considerando un umbral Th, cuyo valor se calcula experimentalmente con base en el estado del arte, a través de la ecuación 2.1 y una muestra de 50 imágenes de $x \times y$ píxeles.

$$Th = \frac{\sum_{i=1}^{x} \sum_{j=1}^{y} I(i, j)}{(x \times y)} \tag{2.1}$$

5. Erosiona B, utilizando un elemento estructurado circular E de radio $R = 5$, considerando que la silueta del cuerpo humano es curvilínea y una estructura circular ayuda a redondear el contorno.

6. Elimina residuos en B utilizando la técnica de 4 vecinos como se muestra en la Figura 2.3, entendiendo como residuo al conjunto de píxeles aislados de la silueta de la persona.

7. Rellena huecos en B, entendiendo que un hueco es un conjunto de píxeles de B con valor 0 al interior del contorno de la silueta de la persona.

8. Calcula la ecuación de la recta de regresión, la cual se usa en el paso 10 para delimitar el cluster donde se detectó movimiento (Ecuación 2.2):

$$y' = a_0 + a_1 x \tag{2.2}$$

Donde y' es la ordenada estimada de los n puntos (x, y), que corresponden a los píxeles en B cuyo valor es 1, mientras que a_0 y a_1 son constantes que obtienen su valor de las ecuaciones normales de la recta de mínimos cuadrados, las cuales en esencia forman un sistema de dos ecuaciones con sus incógnitas cuya solución son las Ecuaciones 2.3 y 2.4.

$$a_0 = \frac{(\sum_{i=1}^{n} y_i)(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} x_i y_i)}{n(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)^2} \tag{2.3}$$

$$a_1 = \frac{n(\sum_{i=1}^{n} x_i y_i) - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)^2} \tag{2.4}$$

9. Calcula los componentes principales considerando $\mathbf{x}$ que representa los n puntos con coordenadas (x, y) que corresponden a los píxeles en B cuyo valor es 1.
El vector media μ está dado por:

$$\mu = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \tag{2.5}$$

De manera similar, la matriz de covarianza $n \times n$, Σ, se puede aproximar por:

$$\Sigma = (\mathbf{x} - \mu)(\mathbf{x} - \mu)^T \tag{2.6}$$

Debido a que Σ es real y asimétrica, siempre es posible encontrar un conjunto de n *eigenvectores ortonormales*. Por lo tanto, la transformada de los componentes principales, también llamada *transformada de Hotelling*, se obtiene por:

$$\mathbf{y} = A(\mathbf{x} - \mu) \tag{2.7}$$

Claramente, los elementos del vector $\mathbf{y}$ no están correlacionados. Por lo tanto, la matriz de *covarianzas* Σ es diagonal. Los renglones de A son los *eigenvectores* $[\vec{v}_1, \vec{v}_2]$ normalizados de Σ, donde los $eigenvalores$, λ_1 y λ_2, con $\lambda_1 \geq \lambda_2$, corresponden a $\vec{v}_1$ y $\vec{v}_2$ respectivamente. Debido a que A es real y simétrica, estos vectores

forman una base ortonormal, y sus matrices inversa y transpuesta son iguales. Por lo tanto, se pueden obtener los valores de $\mathbf{x}$ a través de la transformación inversa.

$$\mathbf{x} = A^T \mathbf{y} + \mu \tag{2.8}$$

10. Delimita la región de movimiento. Para esto, se calculan los puntos extremos: $P1(x_1, y_1)$, $P2(x_2, y_2)$, $P3(x_3, y_3)$ y $P4(x_4, y_4)$ de las rectas L_1 y L_2, cuyos vectores de dirección son los eigenvectores $\vec{v}_1$ y $\vec{v}_2$. Dichas rectas ortogonales se cruzan en el punto medio $O(\bar{x}_i, \bar{y}_i)$ de los píxeles en B cuyo valor es 1. Por consiguiente, como se muestra en la Figura 2.4, los puntos: $E1(x_4, y_1)$, $E2(x_2, y_1)$, $E3(x_2, y_3)$ y $E4(x_4, y_3)$ delimitan la región de movimiento. Para este caso, los puntos que forman la recta L_1 también se pueden calcular utilizando la recta de regresión de la Ecuación 2.2.

2.6.2. Ajuste del modelo de forma activa

Una vez hallada la región de movimiento, donde se asume se encuentra un peatón, los puntos que delimitan esta región: el centroide y sus esquinas, se utilizan como referencia para ajustar un modelo de forma activa (*ASM*). Como se muestra en la Figura 2.5, el ajuste de un modelo de forma activa se realiza en dos fases, un entrenamiento donde se crea el modelo y una fase donde se ajusta este modelo.

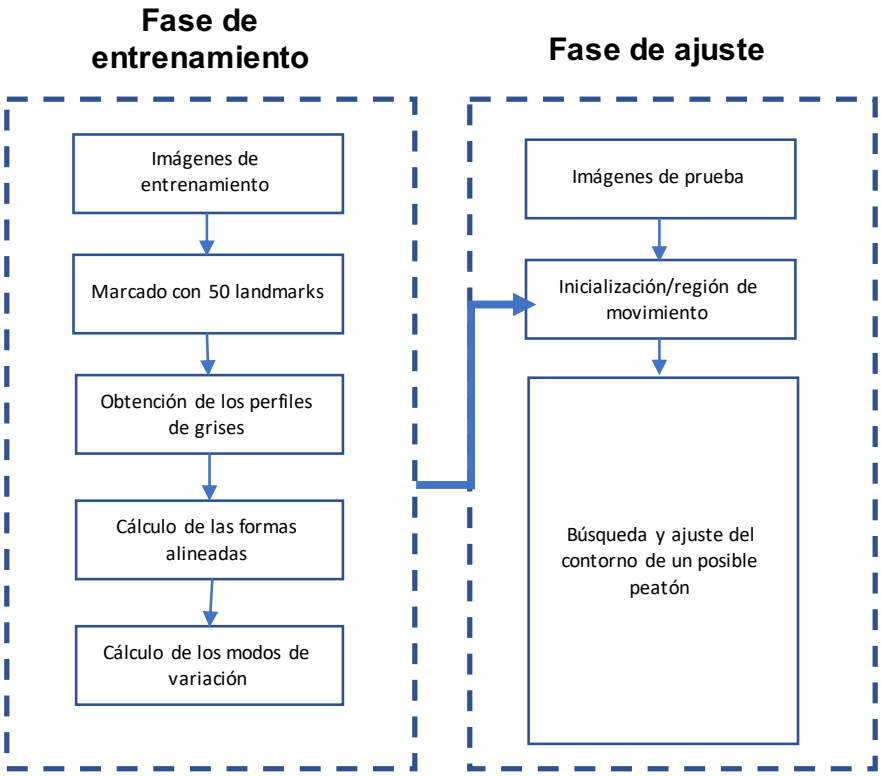


Figura 2.5: Para el ajuste de un modelo de forma activa, se realizan en dos fases: entrenamiento y ajuste.

Fase de entrenamiento

El fin de esta fase es construir un modelo de distribución de puntos *(PDM, por sus siglas en inglés)*, para lo cual, como se muestra en la Figura 2.5 se ejecutan 5 etapas:

1. **Obtener imágenes de entrenamiento**. Para esta investigación se seleccionan escenas de las bases de datos *CDNet 2014* y de *CASIA Gait dataset*. En lo que se refiere a *CDNet 2014* se seleccionan tres videosecuencias con peatones: *office*, *PETS2006* y *sofa*; de las cuales las dos primeras pertenecen a la categoría de *baseline*, mientras que la tercera pertenece a la categoría *intermittent object motion*. Las tres videosecuencias seleccionadas contienen variaciones, principalmente de iluminación y de fondo. Las videograbaciones fueron capturadas en formato *MJPEG*, con una velocidad de 0.17 *frames* por segundo a una resolución de 360×240 píxeles en las escenas *office* y *sofa*, mientras que los frames de la escena *PETS2006* tienen una resolución de 720×576 *píxeles*. En lo que se refiere a la base de datos *CASIA Gait dataset*, contiene videosecuencias donde se almacenan videos codificados con un tamaño de 320×240 *píxeles*; además, contiene información de 124 peatones los cuales caminan en 4 diferentes direcciones: 0^o, 36^o, 54^o y 90^o. Para la experimentación con esta base de dato se seleccionaron 4 escenas, en las cuales 4 sujetos caminan en las 4 direcciones.

2. **Marcado con 50 *landmarks***. Para esto se realiza el marcado de *50* puntos estratégicos alrededor del contorno de un peatón (Figura 2.6), construyendo un conjunto de entrenamiento denotado como:

$$\mathbf{x} = (x_1, y_1, x_2, y_2, ..., x_n, y_n)^T \tag{2.9}$$

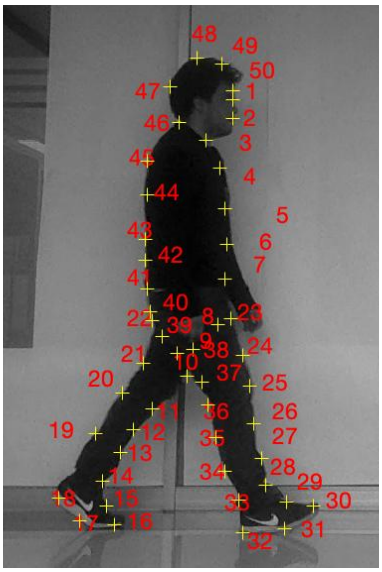


Figura 2.6: Marcado con 50 *landmarks* alrededor del contorno de un peatón.

3. **Obtención de los perfiles de grises**. Por cada uno de los **50** *landmarks* alrededor del contorno del modelo, se colectan los niveles de grises de los píxeles ortogonales. Como se muestra en la Figura 2.7, se obtiene un conjunto de puntos ortonormales en cada *landmark*, considerando *20* píxeles dentro y fuera de la silueta del peatón, respectivamente.

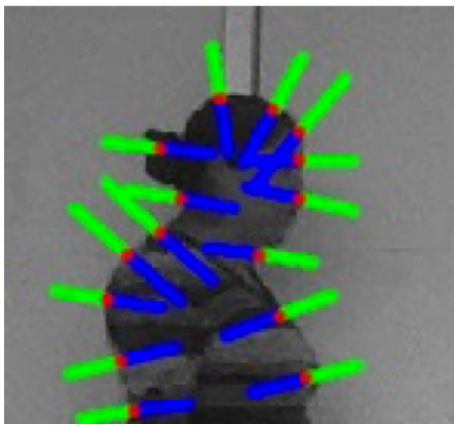


Figura 2.7: Obtención de los perfiles de grises en cada uno de los 50 *landmarks* alrededor del peatón.

4. **Cálculo de las formas alineadas**. La Figura 2.8, muestra el marcado de 50 *landmarks* en imágenes de entrenamiento y escala del peatón contenido en la imagen. Para comparar el par equivalente de coordenadas de las diferentes formas deben estar alineados con respecto a un sistema de referencia. Por lo tanto, se utilizan matrices de transformación geométrica, para aplicar escalado, rotación y traslación de las formas de entrenamiento, a fin de obtener formas alineadas, minimizando el error $\mathbf{E}$, Ecuación 2.10.

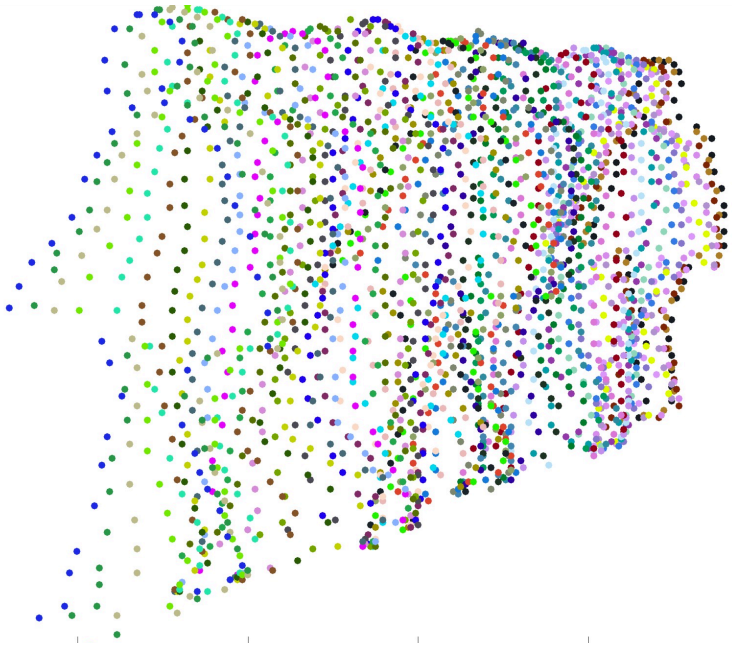


Figura 2.8: Marcado de 50 landmarks en imágenes de entrenamiento.

$$\mathbf{E} = (\mathbf{y} - \mathbf{M}\mathbf{x})^{\mathbf{T}}\mathbf{W}(\mathbf{y} - \mathbf{M}\mathbf{x}), \tag{2.10}$$

donde $\mathbf{W}$ es una matriz diagonal cuyos elementos son factores de ponderación para cada punto de referencia y $\mathbf{M}$ representa la transformación geométrica de la rotación θ, la traslación $\mathbf{t}$ y escalado s. Los factores de pesos se establecen en relación con el desplazamiento entre las posiciones calculadas de los puntos de referencia antiguos y nuevos a lo largo del perfil. Si el desplazamiento es grande, entonces el factor de pesos correspondiente en la matriz se establece bajo; Si el desplazamiento es pequeño, entonces la ponderación con los pesos es alta. Dado un solo punto, denotado por $[\mathbf{x_0}, \mathbf{y_0}]^\mathbf{T}$, la transformación geométrica se define como:

$$M \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = s \begin{pmatrix} cos\theta & sin\theta \\ sin\theta & cos\theta \end{pmatrix} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \tag{2.11}$$

Después de aplicar la transformación geométrica, con los parámetros de pose, $\theta, \mathbf{t}, s$, se obtiene, la proyección de $\mathbf{y}$ en el modelo de coordenadas del *frame*:

$$\mathbf{x_p} = \mathbf{M^{-1}y} \tag{2.12}$$

Finalmente, los parámetros del modelo son actualizados como:

$$\mathbf{b} = \mathbf{\Phi^T}(\mathbf{x_p} - \bar{\mathbf{x}}) \tag{2.13}$$

Como resultado del procedimiento de búsqueda a lo largo de los perfiles, se obtiene el desplazamiento óptimo de un punto de referencia, con lo cual se genera una nueva forma en el marco de coordenadas de la imagen $\mathbf{y}$, el conjunto de entrenamiento alineado se muestra en la Figura 2.9

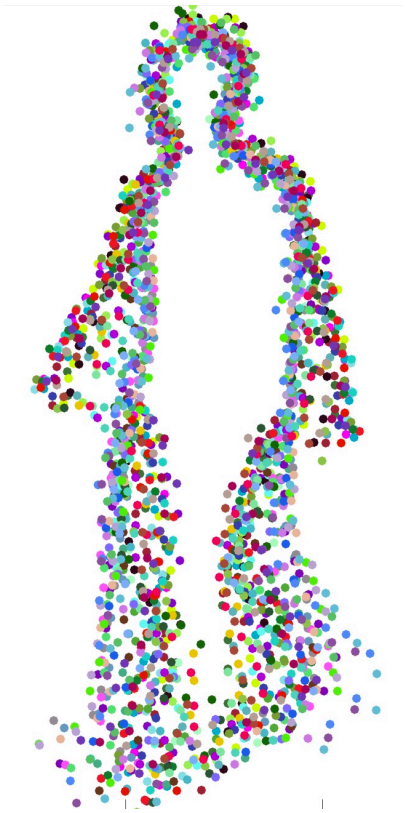


Figura 2.9: Formas alineadas del conjunto de entrenamiento de un peatón.

5. **Cálculo de los modos de variación**. Las coordenadas de los puntos (*landmarks*) del peatón para cada imagen de entrenamiento se almacenan y analizan estadísticamente para extraer las variaciones de forma, considerando que un conjunto de *landmarks* representa la forma del objeto. El conjunto de entrenamiento de esta investigación se compone de 50 formas diferentes, llamado conjunto de entrenamiento. Aunque cada forma en el conjunto de entrenamiento está en el espacio bidimensional, se puede modelar la forma con un número reducido de parámetros utilizando *análisis de componentes principales (PCA)*. Considerando que se tienen m formas en el conjunto de entrenamiento $\mathbf{x_i}$, para $i = 1, .., m$. El *análisis de componentes principales* es el siguiente:

   a) Se calcula la media de las m formas de la muestra en el conjunto de entrenamiento.

$$\bar{x} = \frac{1}{m}\sum_{i=1}^{m} x_i \tag{2.14}$$

   b) Se calcula la matriz de covarianza (**S**) del conjunto de entrenamiento.

$$\mathbf{S} = \frac{1}{m}\sum_{i=1}^{m} \mathbf{dx_i dx_i}^T \tag{2.15}$$

c) Se construye la matriz de *eigenvectores*.

$$\phi = [\phi_1 \mid \phi_2 \mid \cdots \mid \phi_q], \tag{2.16}$$

donde $\phi_j, j = 1, ..., q$ representa los *eigenvectores* de $\mathbf{S}$ correspondiente a los q más largos *eigenvalores*.

d) Dado que ϕ y $\bar{x}$ de cada forma puede ser aproximada como:

$$\mathbf{x_i} \approx \bar{x} + \phi \mathbf{b}_i, \tag{2.17}$$

donde:

$$\mathbf{b_i} = \phi^T(\mathbf{x}_i - \bar{\mathbf{x}}) \tag{2.18}$$

Para efectos de esta investigación se utilizan los q eigenvalores que representen el 98 % de la variación de formas. Se obtienen así formas de peatón aceptables pudiéndose evaluar la distribución de $\mathbf{b}$, para restringir $\mathbf{b}$ a valores aceptables, donde se puede aplicar condiciones duras a cada elemento de $\mathbf{b_i}$ o con restricciones a $\mathbf{b}$ para ser representado en un *hiperelipsoide*.

Fase de ajuste

La fase de ajuste, busca adaptar el modelo de forma a una imagen de prueba, a través de las siguientes etapas:

1. **Imágenes de prueba**. Para llevar a cabo esta investigación dos tipos de conjuntos de datos fueron utilizados, *CDnet2014* y *CASIA Gait dataset* de los cuales se escogieron 50 imágenes para realizar la experimentación con validación cruzada.

2. **Inicialización de la región de movimiento**. En la etapa de ajuste del modelo *ASM*, es necesario obtener las variables de traslación (tx, ty), por lo cual se necesita obtener estas variables de traslación de la técnica *DMSF*, con lo cual se obtienen las coordenadas específicas para que *ASM* pueda realizar la búsqueda de los perfiles de grises y ajustar la silueta del peatón.

3. **Búsqueda y ajuste del contorno de un posible peatón**. Con base a los parámetros de pose y de forma, se selecciona una imagen nueva (en escala de grises) que no pertenece al conjunto de entrenamiento; se busca hacer coincidir esta nueva forma en el conjunto de coordenadas de $\mathbf{x}$ (ver Ecuación 2.9). Para interpretar una forma dada en la imagen de entrada basada en el modelo de forma, se debe encontrar el conjunto de parámetros que mejor se adapte al modelo a la imagen. Suponiendo que el modelo de forma representa límites y bordes fuertes del objeto, un perfil en cada punto de referencia tiene una estructura local similar a un borde.
   Dado que $\mathbf{g}_j$, $j = 1, ..., n$, será la derivada normalizada de un perfil local de longitud K a través del $j - esimo\ landmark$, y $\bar{\mathbf{g}}_j$ y $\mathbf{S}_j$, la media y covarianza correspondientes. El perfil más cercano se considera con la distancia Mahalanobis mínima entre la muestra y la media del modelo:

$$f(\mathbf{g_{j,m}}) = (\mathbf{g_{j,m}} - \bar{\mathbf{g}})^T \mathbf{S}_j^{-1} (\mathbf{g_{j,m}} - \bar{\mathbf{g}_j}) \tag{2.19}$$

   Donde $\mathbf{g_{j,m}}$; $\mathbf{g_j}$ el cual es desplazado por $\mathbf{m}$ muestras a lo largo de la dirección normal del borde de la imagen correspondiente.

2.6.3. Evaluación

Base de datos experimentales

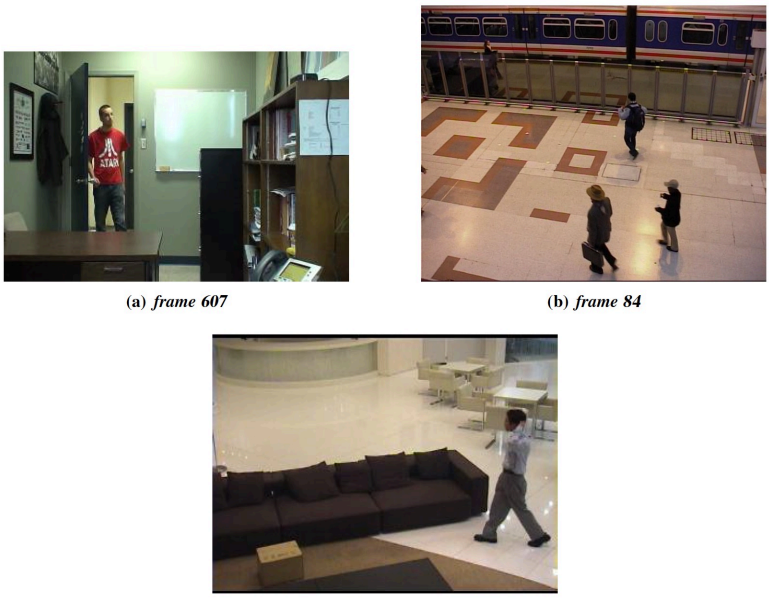
Las bases de datos (*datasets*) de imágenes, son utilizados para proporcionar muestras representativas que contiene información relevante de peatones, estas escenas pueden ser obtenidas desde videocámaras estáticas o en movimiento. Cabe mencionar que las bases de datos contienen peatones etiquetados, además de imágenes *ground truth* para su evaluación y comparación con resultados recientes del estado del arte.
Es por eso que en la literatura mencionan bases de datos que los investigadores están utilizando en la identificación de peatones, tales como: *Caltech-USA*, *Caltech-Japón*, *INRIA*, *ETH*, *TUD-Bruselas* y *Daimler (Daimler-DB)*, las cuales se encuentra disponibles en línea. Además que algunos sitios web permiten el envío de las comparaciones de los resultados obtenidos con nuevos métodos. Los conjuntos de datos *Caltech-USA*, de *INRIA* y *KITTI* son analizados para ver sus diferencias en oclusión, distancia de los peatones, etc, [69, 70, 71, 72].
Sin embargo, para propósito de esta investigación se utilizan las bases de datos *CDnet2014* y *CASIA Gait Dataset* las cuales son citadas en las investigaciones actuales [73, 74, 75], debido a que incluye escenas con un nivel de complejidad alto para validar los algoritmos de detección de movimiento.
CDnet2014, contiene vídeos de escenas cotidianas capturados por una cámara estática y diversas escenas grabadas desde interiores y exteriores. Estos vídeos se han grabado desde cámaras IP de baja resolución, videocámaras de alta resolución, cámaras comerciales *PTZ*, cámaras de infrarrojo, etc. Como consecuencia, las resoluciones espaciales de los videos en el dataset *CDnet2014* varían desde el tamaño de resolución de 320×240 a 720×486 píxeles. *CDnet2014* es un *dataset* con diversas escenas y ambientes, donde aparecen peatones, autos, etc. También ofrece una serie de imágenes de *groundtruht* para su evaluación con resultados binarizados. La base de datos *CDnet 2014*, está compuesta por aproximadamente 90,000 *frames* en 31 videosecuencias representando a las siguientes categorías:

- **Baseline**: Contiene 4 videos con una mezcla de desafíos leves y fáciles de utilizar.
- **Dynamic background**: Contiene 6 vídeos representando escenas al aire libre con fuerte movimiento de fondo.
- **Jitter camera**: Representa 4 vídeos capturados con cámaras móviles.
- **Shadow**: Está compuesto de 6 videos con movimientos suaves y fuertes.
- **Intermittent Object Motion**: Contiene 6 vídeos con escenarios conocidos por causar efectos fantasma, por ejemplo, contiene objetos inmóviles que cambian de posición entre los *frames*.
- **Thermal**: Está compuesto de 5 videos capturados por cámaras infrarojas.

En esta investigación se seleccionaron tres conjuntos de dos categorías de la base de datos *CDnet2014*, los cuales tienen escenarios que tienen un nivel de dificultad alto para detectar movimiento:

a) ***Baseline***: Corresponde a dos videosecuencias que son **office** y **PETS2006** como se aprecia en la Figura 2.10 (a) y (b). La primera escena, contiene 2050 *frames* de 360×240 *píxeles* y corresponde a la aparición de un individuo en una oficina donde un peatón ingresa, toma un libro y vuelve a salir. El escenario PETS2006 contiene 1200 *frames* de 720×576 píxeles, en el que aparecen 6 peatones caminando en diferentes direcciones en una estación de trenes. En este caso incluyen variaciones que afectan la detección: oclusión, reflejos en el piso y partes oscuras.

b) ***Intermittent Object Motion***: De esta categoría se eligió el escenario **sofa** (Figura 2.10 (c)). Éste se compone por 2750 *frames* de 320×240 píxeles, incluyendo variaciones como: iluminación, camuflaje y sombras.

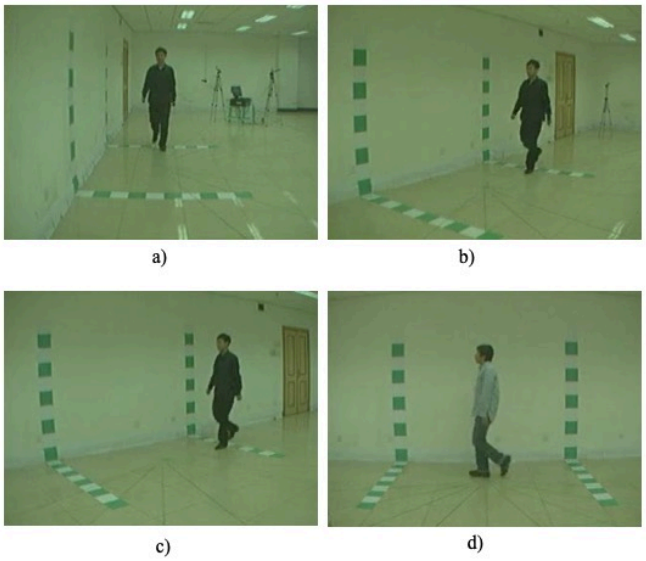


(a) *frame 607* (b) *frame 84*

(c) *frame 630*

Figura 2.10: Escenas originales de las escenas (a) office, (b) PETS2006 y (c) sofa.

En lo que se refiere a la base de datos *CASIA Gait dataset*, el cual contiene videosecuencias donde se almacenan videos con una resolución de 320×240 *píxeles*, además contiene información de 124 peatones los cuales caminan en diferentes direcciones. Para propósito de experimentación, se seleccionaron 4 escenas en las cuales caminan 4 sujetos en cuatro direcciones: 0^o, 36^o, 54^o y 90^o.



a) b)

c) d)

Figura 2.11: Escenas de la base de datos CASIA Gait dataset utilizadas en esta investigación: a) Peatón caminando a 0 grados, b) Peatón caminando a 36 grados, c) Peatón caminando a 54 grados y d) Peatón caminando a 90 grados.

Cálculo del error del ajuste

Para verificar la precisión del ajuste logrado por el modelo de forma activa, se realiza una validación cruzada, la cual es un caso especial de ***k-fold-cross validation***, donde **k** es igual al número de instancias en los datos. Es decir, en cada iteración **k-1** observaciones se utilizan para el entrenamiento y el modelo se prueba con la observación que queda fuera de las **k-ésimas** [76].
Para calcular el error de ajuste se utiliza como métrica la distancia euclidiana (e) entre los puntos (x_g,y_g) del *ground truth* y los puntos ajustados (x_a,y_a) obtenidos por el ajuste del modelo de forma activa.

Entonces, el cálculo del error de ajuste por cada punto del modelo de forma activa esta dado por:

$$e = \sqrt{(x_g - x_a)^2 + (y_g - y_a)^2} \tag{2.20}$$

Y por tanto, el cálculo del error de ajuste medio es:

$$\bar{e} = \frac{1}{K} \sum_{j=1}^{K} e_k \tag{2.21}$$

donde $K = 50$, que corresponde al número de puntos que integran el modelo de forma activa.
Entonces, la gran media de los errores de ajuste de cada una de las 50 iteraciones de la validación cruzada es:

$$\bar{\bar{e}} = \frac{1}{n} \sum_{i=1}^{n} \bar{e}_i \tag{2.22}$$

donde $n = 50$, que corresponde al número de iteraciones de la validación cruzada.
Finalmente, se reporta la gran media de los errores de ajuste de las 50 iteraciones de la validación cruzada y representada gráficamente con diagramas de caja. Considerando que el diagrama de caja es una herramienta útil para analizar el nivel de dispersión de los errores de ajuste medio obtenidos. Se presenta un diagrama de caja por cada uno de los conjuntos de imágenes experimentales de la base de datos *CDnet2014*.
El diagrama de caja es una forma estandarizada que muestra la distribución de datos basada en cuatro parámetros estadísticos: *valor mínimo*, primer cuartil (Q1), mediana, tercer cuartil (Q3) y *valor máximo*. El diagrama de caja también muestra: valores atípicos, simetría de los datos y nivel de sesgo en los datos (ver Figura 2.12). Donde:

- **Media Q2 o percentil 50**: Es el valor medio del conjunto de datos.
- **Primer cuartil (Q1 o percentil 25)**: Es el número medio entre el valor *mínimo* ($Q3 + 1.5IQR$) y la mediana del conjunto de datos.
- **Tercer cuartil (Q3 o percentil 75)**: Es el valor medio entre la mediana y el valor *máximo* ($Q1 - 1.5IQR$) del conjunto de datos.
- **Rango intercuartil (IQR)**: Percentil 25 al 75.

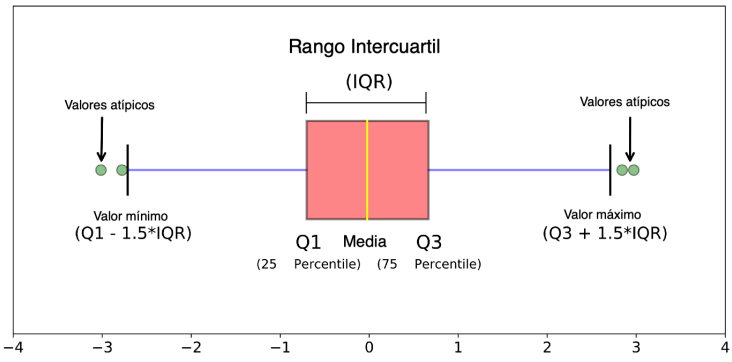


Figura 2.12: Partes que integran un *box plot*. En esta gráfica se muestran los bigotes como líneas azules y los valores atípicos como puntos verdes.

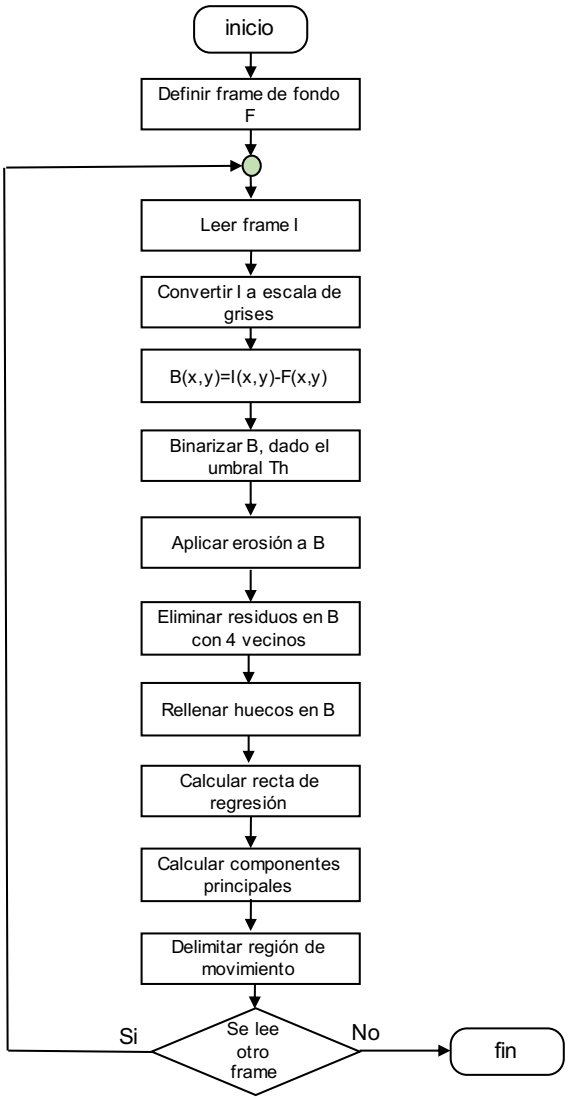


Figura 2.2: Diagrama de flujo, donde se muestran las etapas para detectar una región de movimiento con la técnica DMSF.

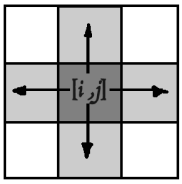


Figura 2.3: Eliminación de *píxeles* cercanos con la técnica de 4 vecinos $[i+1, j], [i-1, j], [i, j+1], [i, j-1]$.

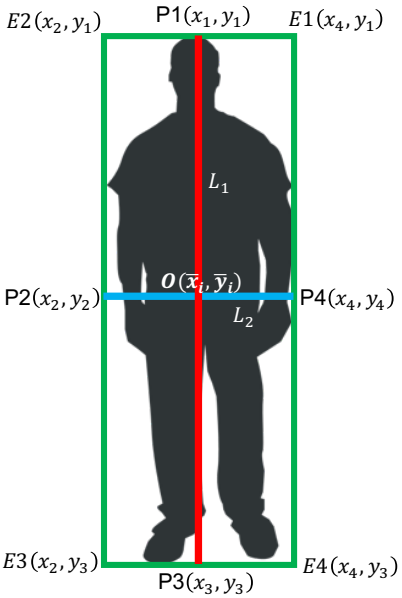


Figura 2.4: Región de movimiento (rectángulo verde) delimitado por los puntos extremos: $E1(x_4, y_1)$, $E2(x_2, y_1)$, $E3(x_2, y_3)$ y $E4(x_4, y_3)$, de las rectas L_1 y L_2, cuyos vectores de dirección son los eigenvectores $\vec{v}_1$ y $\vec{v}_2$.

Bibliografía

[1] X. Yongzhe, "Smart surveillance camera based on pattern," 2014. 1, 3

[2] X. Wang, "Intelligent multi-camera video surveillance: A review," *Pattern recognition letters*, vol. 34, no. 1, pp. 3–19, 2013. 1

[3] L. Seidenari and M. Bertini, "Non-parametric anomaly detection exploiting space-time features," in *Proceedings of the 18th ACM international conference on Multimedia*. ACM, 2010, pp. 1139–1142. 1, 3

[4] S. R. Valdehita, E. M. D. Ramiro, J. M. García, and L. L. Moreno, "Carga mental en vigilantes de seguridad: diferencias por sexo y capacidad atencional," *EduPsykhé: Revista de psicología y psicopedagogía*, vol. 7, no. 2, pp. 213–230, 2008. 1, 3

[5] M. J. M. Vasconcelos and J. M. R. Tavares, "Human motion segmentation using active shape models," in *Computational and Experimental Biomedical Sciences: Methods and Applications*. Springer, 2015, pp. 237–246. 2, 7

[6] P. K. Mishra and G. Saroha, "A study on classification for static and moving object in video surveillance system," *International Journal of Image, Graphics and Signal Processing (IJIGSP)*, vol. 8, no. 5, pp. 76–82, 2016. 3

[7] E. Wallace, C. Diffley, and J. Aldridge, "Good practice for the management and operation of town centre cctv," 1997. 3

[8] X. Wang, M. Wang, and W. Li, "Scene-specific pedestrian detection for static video surveillance," *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, no. 2, pp. 361–374, 2014. 3

[9] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, "Performance measures and a data set for multi-target, multi-camera tracking," in *European Conference on Computer Vision*. Springer, 2016, pp. 17–35. 3

[10] N. Arteaga Botello, "Regulación de la videovigilancia en méxico. gestión de la ciudadanía y acceso a la ciudad," *Espiral (Guadalajara)*, vol. 23, no. 66, pp. 193–238, 2016. 3

[11] ——, "Video-vigilancia del espacio urbano: tránsito, seguridad y control social," *Andamios*, vol. 7, no. 14, pp. 263–286, 2010. 4

[12] V. M. Sánchez Valdés, "¿ son efectivas las cámaras de video vigilancia para reducirlos delitos?" 2016. 4

[13] M. Alzantot and M. Youssef, "Uptime: Ubiquitous pedestrian tracking using mobile phones," in *Wireless Communications and Networking Conference (WCNC), 2012 IEEE*. IEEE, 2012, pp. 3204–3209. 4

[14] J. R. Van Huis, H. Bouma, J. Baan, G. J. Burghouts, P. T. Eendebak, R. J. den Hollander, J. Dijk, and J. H. van Rest, "Track-based event recognition in a realistic crowded environment," in *Optics and Photonics for Counterterrorism, Crime Fighting, and Defence X; and Optical Materials and Biomaterials in Security and Defence Systems Technology XI*, vol. 9253. International Society for Optics and Photonics, 2014, p. 92530E. 4

[15] D. Ramanan, D. A. Forsyth, and A. Zisserman, "Tracking people by learning their appearance," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 1, pp. 65–81, 2007. 4

[16] X. Yang, S. Yuan, and Y. Tian, "Assistive clothing pattern recognition for visually impaired people," *IEEE Transactions on Human-Machine Systems*, vol. 44, no. 2, pp. 234–243, 2014. 4

[17] C. Lassner, G. Pons-Moll, and P. V. Gehler, "A generative model of people in clothing," *arXiv preprint arXiv:1705.04098*, 2017. 4

[18] N. Jetchev and U. Bergmann, "The conditional analogy gan: Swapping fashion articles on people images," in *Computer Vision Workshop (ICCVW), 2017 IEEE International Conference on*. IEEE, 2017, pp. 2287–2292. 4

[19] L. Ma, X. Jia, Q. Sun, B. Schiele, T. Tuytelaars, and L. Van Gool, "Pose guided person image generation," in *Advances in Neural Information Processing Systems*, 2017, pp. 405–415. 4

[20] S. Hamdi, H. Faiedh, C. Souani, and K. Besbes, "A lighting independent vision based system for driver assistance," in *Design & Test Symposium (IDT), 2016 11th International*. IEEE, 2016, pp. 328–333. 4

[21] S. K. Kwon, E. Hyun, J.-H. Lee, J. Lee, and S. H. Son, "A low-complexity scheme for partially occluded pedestrian detection using lidar-radar sensor fusion," in *Embedded and Real-Time Computing Systems and Applications (RTCSA), 2016 IEEE 22nd International Conference on*. IEEE, 2016, pp. 104–104. 4

[22] Y. Tian, P. Luo, X. Wang, and X. Tang, "Deep learning strong parts for pedestrian detection," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1904–1912. 4

[23] J. Hosang, M. Omran, R. Benenson, and B. Schiele, "Taking a deeper look at pedestrians," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4073–4082. 4

[24] M. Mathias, R. Benenson, R. Timofte, and L. Van Gool, "Handling occlusions with franken-classifiers," in *Computer Vision (ICCV), 2013 IEEE International Conference on*. IEEE, 2013, pp. 1505–1512. 4

[25] C. Zhu and Y. Peng, "A boosted multi-task model for pedestrian detection with occlusion handling," *IEEE transactions on image processing*, vol. 24, no. 12, pp. 5619–5629, 2015. 4

[26] B. Peralta, L. Parra, and L. Caro, "Evaluation of stacked autoencoders for pedestrian detection," in *Computer Science Society (SCCC), 2016 35th International Conference of the Chilean*. IEEE, 2016, pp. 1–7. 4

[27] H. Zhang, C. Reardon, and L. E. Parker, "Real-time multiple human perception with color-depth cameras on a mobile robot," *IEEE Transactions on Cybernetics*, vol. 43, no. 5, pp. 1429–1441, 2013. 4

[28] C. Sukanya, R. Gokul, and V. Paul, "A survey on object recognition methods," *International Journal of Science, Engineering and Computer Technology*, vol. 6, no. 1, p. 48, 2016. 4

[29] B. Zhou, X. Wang, and X. Tang, "Understanding collective crowd behaviors: Learning a mixture model of dynamic pedestrian-agents," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2871–2878. 4

[30] I. S. Kim, H. S. Choi, K. M. Yi, J. Y. Choi, and S. G. Kong, "Intelligent visual surveillance—a survey," *International Journal of Control, Automation and Systems*, vol. 8, no. 5, pp. 926–939, 2010. 5

[31] M. Paul, S. M. Haque, and S. Chakraborty, "Human detection in surveillance videos and its applications-a review," *EURASIP Journal on Advances in Signal Processing*, vol. 2013, no. 1, p. 176, 2013. 5

[32] W. Hu, T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 34, no. 3, pp. 334–352, 2004. 5

[33] M. Piccardi, "Background subtraction techniques: a review," in *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583)*, vol. 4. IEEE, 2004, pp. 3099–3104. 5

[34] D. A. Migliore, M. Matteucci, and M. Naccari, "A revaluation of frame difference in fast and robust motion detection," in *Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks*. ACM, 2006, pp. 215–218. 5

[35] S. Yalamanchili, W. N. Martin, and J. K. Aggarwal, "Extraction of moving object descriptions via differencing," *Computer graphics and image processing*, vol. 18, no. 2, pp. 188–201, 1982. 5

[36] R. Jain, W. Martin, and J. Aggarwal, "Segmentation through the detection of changes due to motion," *Computer Graphics and Image Processing*, vol. 11, no. 1, pp. 13–34, 1979. 5

[37] B. D. Lucas, T. Kanade *et al.*, "An iterative image registration technique with an application to stereo vision," 1981. 5

[38] B. K. Horn and B. G. Schunck, ""determining optical flow": A retrospective," 1993. 5

[39] T. Bouwmans, "Recent advanced statistical background modeling for foreground detection-a systematic survey," *Recent Patents on Computer Science*, vol. 4, no. 3, pp. 147–176, 2011. 5

[40] W.-x. Kang, W.-z. Lai, and X.-b. Meng, "An adaptive background reconstruction algorithm based on inertial filtering," *Optoelectronics Letters*, vol. 5, no. 6, pp. 468–471, 2009. 5

[41] S. Jiang and Y. Zhao, "Background extraction algorithm base on partition weighed histogram," in *2012 3rd IEEE International Conference on Network Infrastructure and Digital Content*. IEEE, 2012, pp. 433–437. 5

[42] L. Maddalena and A. Petrosino, "A self-organizing approach to background subtraction for visual surveillance applications," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1168–1177, 2008. 5

[43] B. Antic, V. Crnojevic, and D. Culibrk, "Efficient wavelet based detection of moving objects," in *2009 16th International Conference on Digital Signal Processing*. IEEE, 2009, pp. 1–6. 5

[44] S. Messelodi, C. M. Modena, N. Segata, and M. Zanin, "A kalman filter based background updating algorithm robust to sharp illumination changes," in *International Conference on Image Analysis and Processing*. Springer, 2005, pp. 163–170. 5

[45] J. Zheng, Y. Wang, N. L. Nihan, and M. E. Hallenbeck, "Extracting roadway background image: Mode-based approach," *Transportation research record*, vol. 1944, no. 1, pp. 82–88, 2006. 5

[46] P. Ramya and R. Rajeswari, "A modified frame difference method using correlation coefficient for background subtraction," *Procedia Computer Science*, vol. 93, pp. 478–485, 2016. 5

[47] Z. Xu, B. Min, and R. C. Cheung, "A robust background initialization algorithm with superpixel motion detection," *Signal Processing: Image Communication*, vol. 71, pp. 1–12, 2019. 5

[48] J. Yin, L. Liu, H. Li, and Q. Liu, "The infrared moving object detection and security detection related algorithms based on w4 and frame difference," *Infrared Physics & Technology*, vol. 77, pp. 302–315, 2016. 5

[49] S.-h. Lee, G.-c. Lee, J. Yoo, and S. Kwon, "Wisenetmd: Motion detection using dynamic background region analysis," *Symmetry*, vol. 11, no. 5, p. 621, 2019. 6

[50] P. Gupta, Y. Singh, and M. Gupt, "Moving object detecti on using frame difference, background subtraction and sobs for vi deo surveillance application," *pp151-156*, 2014. 6

[51] K. Sehairi, C. Fatima, and J. Meunier, "A benchmark of motion detection algorithms for static camera: Application on cdnet 2012 dataset," in *International Conference on Computer Science and its Applications*. Springer, 2018, pp. 235–245. 6

[52] I. Haritaoglu, D. Harwood, and L. S. Davis, "W/sup 4: real-time surveillance of people and their activities," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 8, pp. 809–830, 2000. 6

[53] F. Marqués and V. Vilaplana, "Face segmentation and tracking based on connected operators and partition projection," *Pattern Recognition*, vol. 35, no. 3, pp. 601–614, 2002. 6

[54] R. Plänkers and P. Fua, "Tracking and modeling people in video sequences," *Computer Vision and Image Understanding*, vol. 81, no. 3, pp. 285–302, 2001. 6

[55] A. Baumberg and D. Hogg, "An efficient method for contour tracking using active shape models," in *Motion of Non-Rigid and Articulated Objects, 1994., Proceedings of the 1994 IEEE Workshop on*. IEEE, 1994, pp. 194–199. 6

[56] D. Kim, S. Lee, and J. Paik, "Active shape model-based gait recognition using infrared images," in *Signal Processing, Image Processing and Pattern Recognition*. Springer, 2009, pp. 275–281. 6

[57] D. Kim and J. Paik, "Gait recognition using active shape model and motion prediction," *IET Computer Vision*, vol. 4, no. 1, pp. 25–36, 2010. 6

[58] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer, "The humanid gait challenge problem: Data sets, performance, and analysis," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 2, pp. 162–177, 2005. 6

[59] H. Sadoghi Yazdi, H. J. Fariman, and J. Roohi, "Gait recognition based on invariant leg classification using a neuro-fuzzy algorithm as the fusion method," *ISRN Artificial Intelligence*, vol. 2012, 2011. 7

[60] S. Ye, S.-P. Lu, and A. Munteanu, "Color correction for large-baseline multiview video," *Signal Processing: Image Communication*, vol. 53, pp. 40–50, 2017. 7

[61] D. Lee and S. Choi, "Multisensor fusion-based object detection and tracking using active shape model," in *Digital Information Management (ICDIM), 2011 Sixth International Conference on*. IEEE, 2011, pp. 108–114. 7

[62] V. Moskvina and A. Zhigljavsky, "Application of the singular spectrum analysis for change-point detection in time series," *Journal of Time Series Analysis, submitted*, 2001. 7

[63] C. Jang and K. Jung, "Human pose estimation using active shape models," *Proceedings of World Academy of Science: Engineering & Technology*, vol. 46, 2008. 7

[64] A. Koschan, S. Kang, J. Paik, B. Abidi, and M. Abidi, "Color active shape models for tracking non-rigid objects," *Pattern Recognition Letters*, vol. 24, no. 11, pp. 1751–1765, 2003. 7

[65] S. Arjun, "Active shape model based pose estimation using hausdorff matching," *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, vol. 3, pp. 411–418, jun 2015. 7

[66] S. Zheng, J. Zhang, K. Huang, R. He, and T. Tan, "Robust view transformation model for gait recognition," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*. IEEE, 2011, pp. 2073–2076. 7

[67] J. Ma and F. Ren, "Detect and track the dynamic deformation human body with the active shape model modified by motion vectors," in *Cloud Computing and Intelligence Systems (CCIS), 2011 IEEE International Conference on*. IEEE, 2011, pp. 587–591. 7

[68] E. Pourjam, I. Ide, D. Deguchi, and H. Murase, "Segmentation of human instances using grab-cut and active shape model feedback." in *MVA*, 2013, pp. 77–80. 8

[69] P. Balasubramanian, S. Pathak, and A. Mittal, "Improving Gradient Histogram Based Descriptors for Pedestrian Detection in Datasets with Large Variations," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2016. 16

[70] H. K. Kim and D. Kim, "Robust pedestrian detection under deformation using simple boosted features," *Image and Vision Computing*, 2017. 16

[71] W. Ouyang and X. Wang, "A discriminative deep model for pedestrian detection with occlusion handling," *CVPR'12*, 2012. 16

[72] R. N. Rajaram, E. Ohn-Bar, and M. M. Trivedi, "An Exploration of Why and When Pedestrian Detection Fails," in *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 2015. 16

[73] A. Miron and A. Badii, "Change detection based on graph cuts," in *2015 International Conference on Systems, Signals and Image Processing (IWSSIP)*. IEEE, 2015, pp. 273–276. 16

[74] M. De Gregorio and M. Giordano, "Change detection with weightless neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2014, pp. 403–407. 16

[75] ——, "Wisardrp for change detection in video sequences." in *ESANN*, 2017. 16

[76] P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-validation," in *Encyclopedia of database systems*. Springer, 2009, pp. 532–538. 18

CAPÍTULO 3

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

En este capítulo se presenta un artículo aceptado para su publicación en *la revista CIENCIA ergo-sum (**ISSN: 2395-8782**). Revista Científica Multidisciplinaria de Prospectiva de la Universidad del Estado de México*, la revista cuenta con una indización Open Journal Systems con los siguientes indíces: biblat, C.I.R.C EC3metrics, Clarivate Analytics, Clase, Dialnet, CONACYT, DOAJ, ERIHPLUS, ibss, iresie, MIAR, latindex, PKP INDEX, redalyc.org, REDIB, SHERPA RoMEO, SSOAR Social science y WorldCat.

CIENCIA *ergo-sum*
Universidad Autónoma del Estado de México
ciencia.ergosum@yahoo.com.mx
E-ISSN: 2395-8782

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

Antonio, Juan Alberto; Romero, Marcelo
Detección de peatones con variaciones de forma al caminar con Modelos de Forma Activa
CIENCIA *ergo-sum*, vol. 27, núm. 3, noviembre 2020-febrero 2021 | **e100**
Universidad Autónoma del Estado de México, México

Antonio, J. A. y Romero, M. (2020). Detección de peatones con variaciones de forma al caminar con Modelos de Forma Activa. CIENCIA *ergo-sum*, *27*(3). http://doi.org/10.30878/ces.v27n3a10

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
mromeroh@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 los resultados evaluados con *leave one out* en cada imagen de 1 080 × 720 pixeles y con la métrica del error cuadrático medio (MSE) se obtiene un promedio total de 12.7 pixeles en la distancia de error entre los *landmarks* originales y los *landmark*s 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 × 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 *Staticals Shape Models* (*SSM*), *Active Contour Models* (*ACM*, *Snakes*), *Active Apearence Models* (*AAM*), etc., (Hill, Thornham y Taylor, 2013; Blake, Curwen y Zisserman, 1993; Enzweiler y Gavrila, 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 Gavrila, 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 caminando, lo que permite analizar la cadencia al caminar (*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 aún 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 en base a 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)performance of the proposed algorithm has been validated by using the HumanID Gait Challenge data set (HGCD. Detectan humanos automáticamente con ASM utilizando 32 *landmarks* para la silueta de un humano a partir de imágenes de 720 x 480 pixeles. 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 seguir 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 pixeles fueron obtenidos de una cámara *RGB* omnidireccional. Al final muestran un 94% de éxito en la detección de peatones.

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 pixeles. 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 a la detección de transeúntes, no mencionan como 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 detecccción 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.

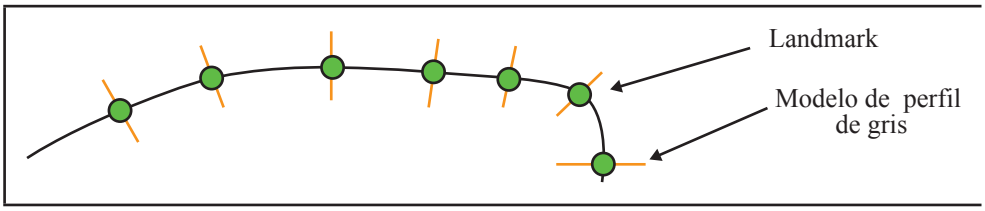


FIGURA 1
Representación del modelado de puntos (landmarks, color verde) y la obtención de los perfiles de grises (línea amarilla)
Fuente: Scott, Cootes y Taylor (2003).

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:

$$x = [x_1, \ldots, x_n, y_1, \ldots, y_n]^T \tag{1}$$

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.

$$X = M(s, \theta)[x] + x \tag{2}$$

donde:

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

$M(s, \theta)[x]$ 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 entramiento alineado (X_i, ecuación 4)

$$X_i = Mx + (tx, ty) \tag{4}$$

c) Obtención de los componentes principales (PCA)

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

$$\bar{x} = \frac{1}{m} \sum_{i=1}^{m} X_i \tag{5}$$

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

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

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

$$\Phi = [\phi_1 | \phi_2 | \ldots | \phi_q] \tag{7}$$

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

$$x_i \approx \bar{x} + \Phi b_i \tag{8}$$

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, el nivel 1 es una imagen con la mitad del número de pixeles 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.

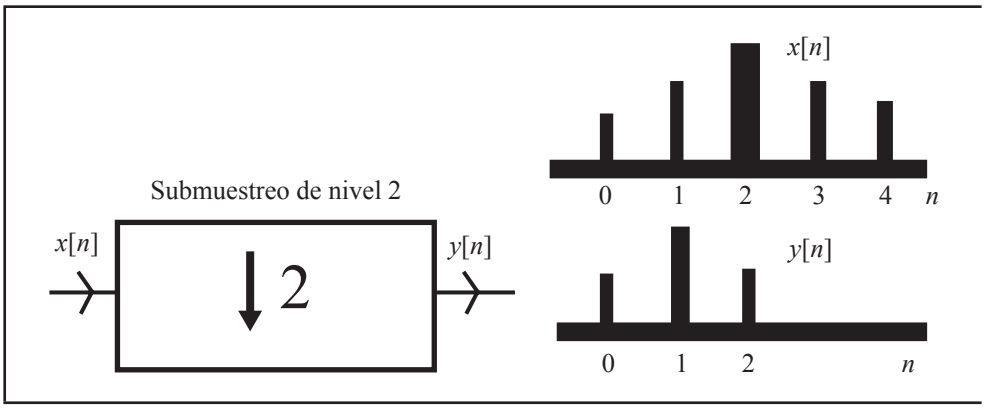


FIGURA 2
Ejemplo de submuestreo en el cambio de resolución en una pirámide de resolución de una imagen
Fuente:Scott, Cootes y Taylor (2003).

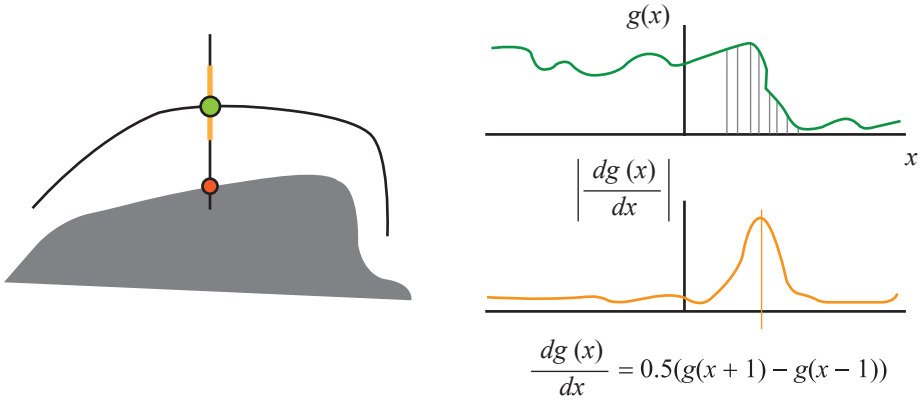
$$y[n] = \sum_n x[2n] e^{-i\omega n} \tag{11}$$

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:

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

$$g_{ajustado} = g + P_g\, b_g \tag{13}$$

$g_{ajustado}$ 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 brevemente 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*) juste (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 pude 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*.

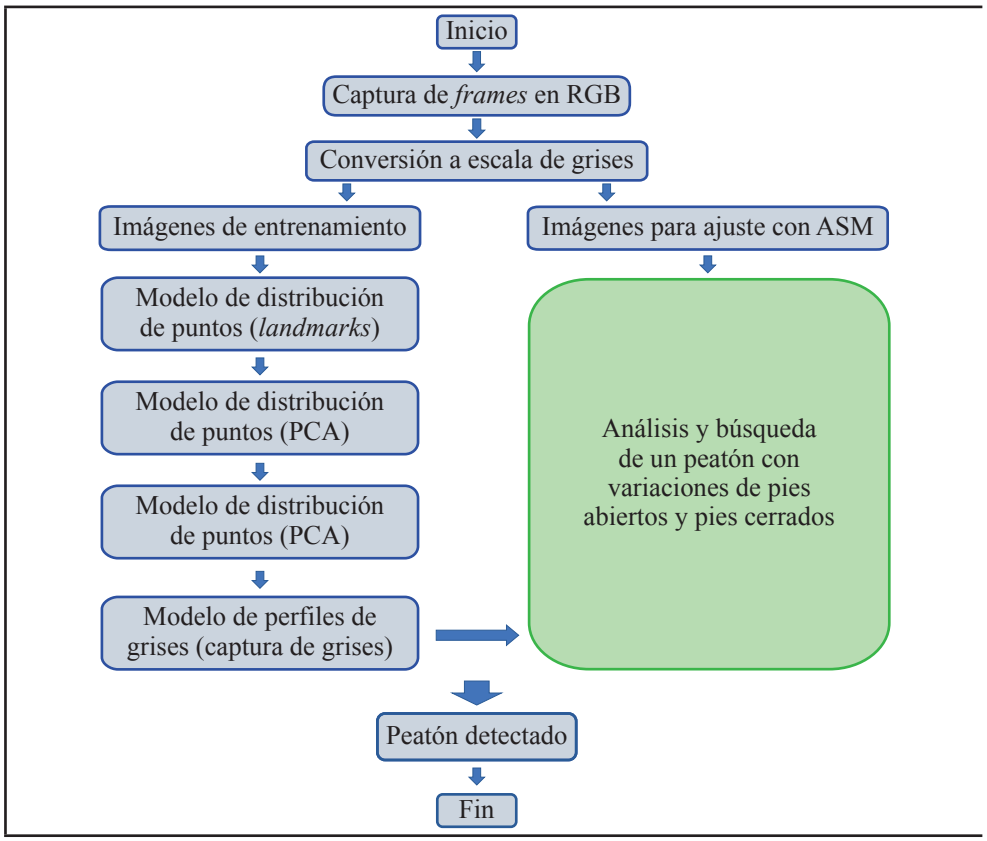


FIGURA 3

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, ... Conjunto Entrenamiento do
4:     Marcar cada imagen con los landmarks predefinidos
5:     Ec. (1)
6: end for
7: for iteration = 1, 2, ... Total fremes do
8:     Cálculo de Componentes Principales
9:     {Ec. 2-9}
10: end for
11: for iteration = 1, 2, ... Total fremes 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 × 3.80 m de ancho en el interior del edificio de posgrado de la Facultad de Ingeniería de la Universidad Autonóma 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 × 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.

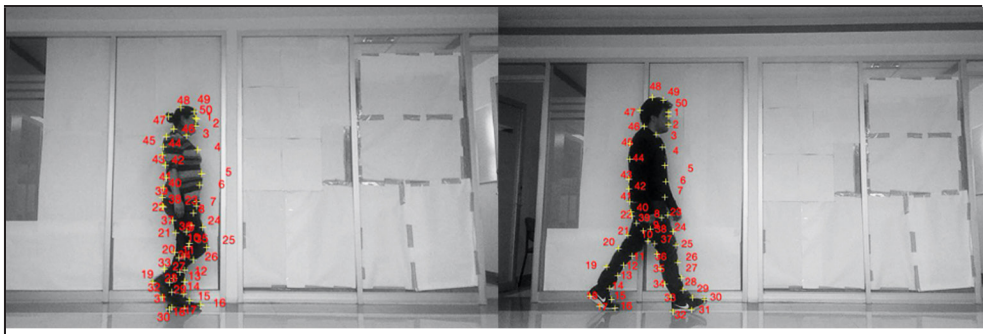
TABLA 1
Extracción de imágenes experimentales en peatón 1 y peatón 2

Cantidad de *frames*	Peatón 1		Peatón 2	
	Pies cerrados	Pies abiertos	Pies cerrados	Pies abiertos
Frames individuales	25	39	44	15
Frames totales	66		71	

Fuente: elaboración propia.

4. 2. Modelado de distribución de puntos

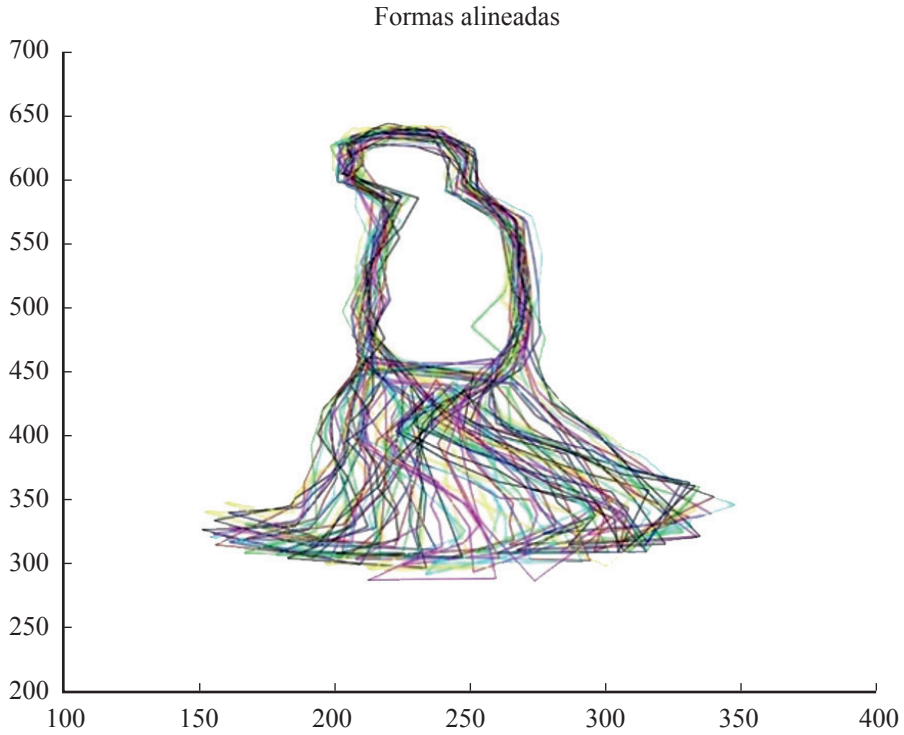
En la etapa de modelado de distribución de puntos se propuso un marcado de 50 *landmarks* alrededor del contorno del peatón (Godil, 2007) tomando en cuenta algunos puntos antropométricos según el modelo CAESAR (Ressler, 2001). De este modelo se pueden identificar los puntos específicos que corresponden 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).



a) Pies cerrados *b*) Pies abiertos

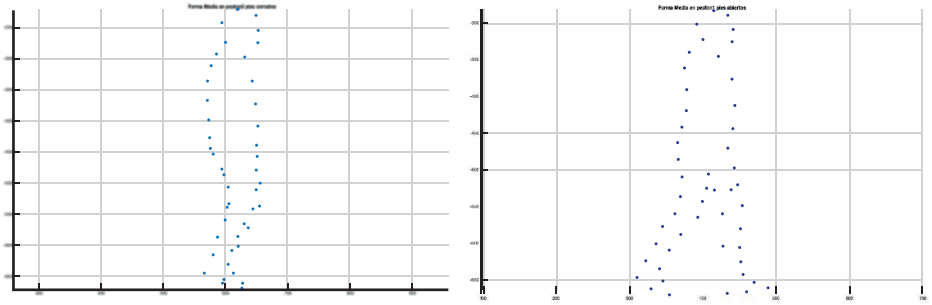
FIGURA 5
Etapas para la detección de un peatón con variaciones de forma con ASM
Fuente: elaboración propia.

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 pixeles de ancho (20 pixeles hacia arriba y 20 pixeles 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.



a) peatón 2, pies cerrados *b*) peatón 1, pies abiertos

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 6*b*) 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.

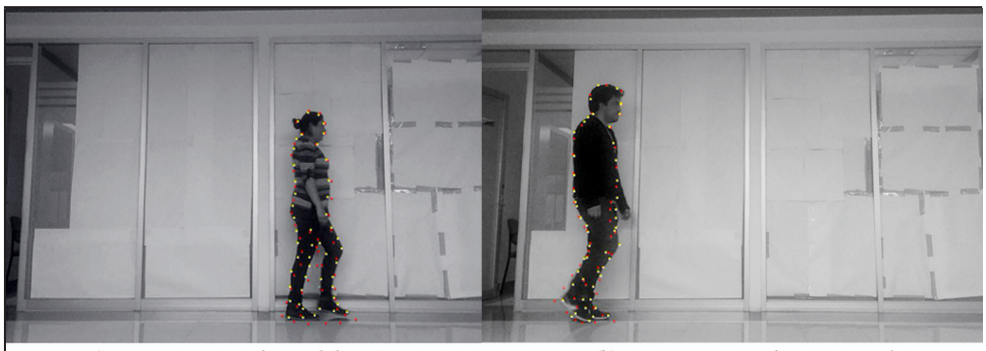


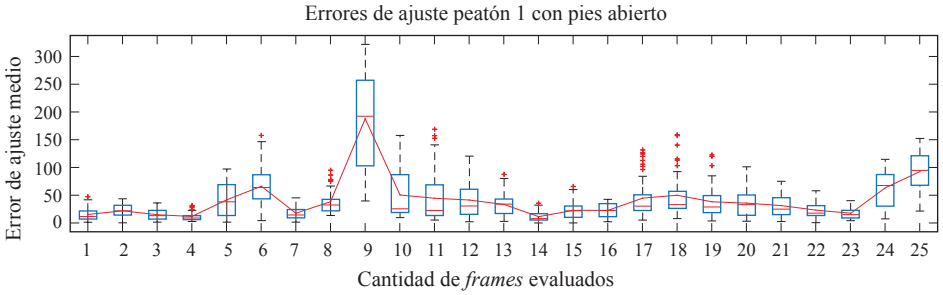
FIGURA 6
Representación del marcado original (puntos amarillos) y el marcado resultante del ajuste (puntos rojos)
Fuente: elaboración propia.

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).

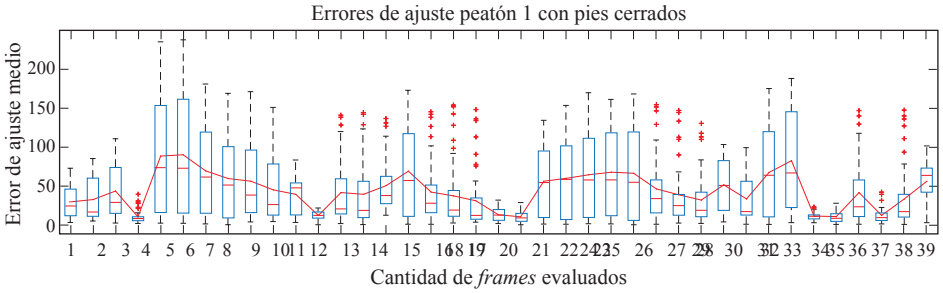
La gráfica 4 muestra el MSE del ajuste del ASM con los 25 *frames* que contienen la escena de peatón 1 (con pies abiertos). Se observa que el peor resultado en el cálculo de la distancia con MSE entre los *landmarks* originales y los *landmarks* ajustados se obtiene con el *frame* 9 a una distancia de 150 pixeles, mientras que el ajuste más equilibrado se encuentra en el peatón 1, donde se observa que el *frame* 4 tuvo un resultado de 10 pixeles. 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 pixeles, 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.



GRÁFICA 4
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.

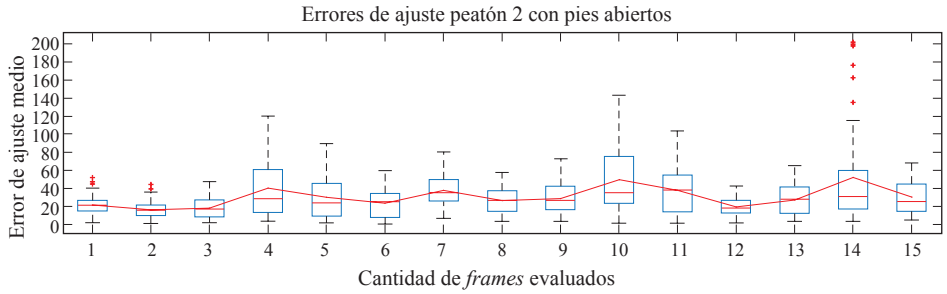
Por su parte, la gráfica 5 corresponde al conjunto de 39 imágenes experimentales del peatón 1 con pies cerrados muestra mediante una línea roja que la distancia promedio de MSE es de 50 pixeles entre los *landmarks* originales y los *landmarks* resultado del ajuste. En algunos *frames* como 4, 12, 34,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



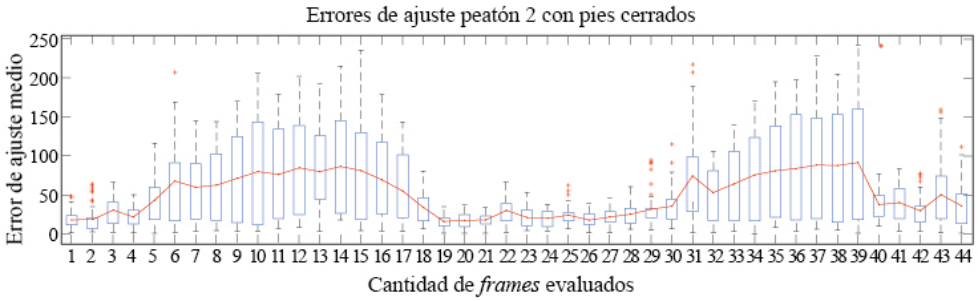
GRÁFICA 5
Representación gráfica de diagrama de caja, donde se muestra la evaluación promedio del MSE en cada uno de los 39 *frames* del peatón 1 con pies cerrados
Nota: elaboración propia.

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 pixeles 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 pixeles 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 pixeles, 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 pixeles).



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 15 *frames* del peatón 2 con pies abiertos
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 pixeles del cálculo de la distancia MSE y con pies abiertos y un conjunto de 25 *frames* tuvo un promedio total de 12.7 pixeles 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 pixeles 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 pixeles. 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 pixeles, 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 *frames*	Error medio de ajuste total	Número de *frames*	Error medio de ajuste total
Pies abiertos	25	41.6	15	30.4
	Número de *frames*	Error medio de ajuste total	Número de *frames*	Error medio de ajuste total
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 los 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 pixeles 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 pixeles 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.

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. y 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. *Procedings 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

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

CAPÍTULO 4

Pedestrian's localization into a video sequence using motion detection and active shape models

En este capítulo se presenta un artículo enviado para su revisión y posible publicación en la revista *IEEE Latin American Transactions (**ISSN:1548-0992**)*, la cual cuenta con una indización en Scopus, Science Citation Index Expanded, Aerospace Database, Civil Engineering Abstracts, Compendex, INSPEC, Metadex y Comunication Abstracts.

IV

Ilse Cervantes via <noreply@latamt.ieeer9.org>
Mié 11/11/2020 12:27 AM
Para: Juan Alberto Antonio Velazquez

Juan Alberto Antonio Velázquez:

Thank you for submitting the manuscript, "Pedestrian's localization into a video sequence using motion detection and active shape models: Localización de peatones en una secuencia de video mediante detección de movimiento y modelos de formas activas" to IEEE Latin America Transactions. With the online journal management system that we are using, you will be able to track its progress through the editorial process by logging in to the journal web site:

Submission URL: https://latamt.ieeer9.org/index.php/transactions/authorDashboard/submission/5123
Username: jaantoniovelazquez

If you have any questions, please contact me. Thank you for considering this journal as a venue for your work.

Ilse Cervantes

Pedestrian's localization into a video sequence using motion detection and active shape models

J.A Antonio, M. Romero, and R. Alejo

Abstract*— Problems of background, scale, low contrast and resolution variations in video sequences, make pedestrian's detection a challenging task in research. For years, different algorithms are aimed to detect moving pedestrians despite variations on background, pose and scale. In this paper we introduce a novel pedestrian's detection method, which is based on two techniques: motion detection using background subtraction (DMSF) and active shape models (ASM). Our proposed active shape model consists of 50 landmarks around a pedestrian's silhouette and a gray scale of 40 pixels per landmark. This research is novel because we are combining a motion detection technique to identify regions of interest (ROI), which might contain a pedestrian, after that, an active shape model is adjusted on every ROI. Our approach is able to locate every pedestrian within the scene and it is robust to pose variations. We are evaluating our approach using cross validation on two state of the art databases: CDnet2014 and CASIA Gait dataset. Ground true data is collected on every experimental video for performance evaluation. We are computing adjustment errors by calculating the Euclidean distance between our ground truth and ASM estimated landmarks. Our best adjustment mean error is achieved on the CASIA Gait dataset, scoring a grand mean of 4.5 pixels.

***Index Terms*— ASM, ROI, landmarks, motion detection.**

I. INTRODUCCIÓN

La detección de personas es de importancia para el estudio de aplicaciones como los sistemas de videovigilancia [1], la cual consiste en detectar sujetos dentro de una imagen digital en 2D o videosecuencia.

Sus desafíos se deben a la variabilidad en la apariencia de los peatones debido a la pose, vestimenta, oclusión, niveles de iluminación, fondo y condiciones climáticas [2–6]. La detección de peatones utilizando cámaras de videovigilancia, necesita de métodos que ayuden a capturar sus diferentes posturas al caminar, las cuales varían a lo largo de la videosecuencia [6]. En la actualidad la detección de peatones se ha realizado a través de diferentes algoritmos, por ejemplo el uso de técnicas deformables como: *staticals shape models* (*SSM*), *active contour models* (*ACM snakes*), y *active apearence models* (*AAM*) entre otros [8–14].

J.A Antonio Universidad Autónoma Del Estado De México, Facultad de Ingeniería, Toluca Estado de México, email: jaantoniov@uaemex.mx.

M. Romero Universidad Autónoma Del Estado De México, Facultad de Ingeniería, Toluca Estado de México, email: mromerohg@uaemex.mx.

R. Alejo Instituto Tecnológico de Toluca, División de Estudios de Posgrado e Investigación email: ralejoe@toluca.tecnm.mx.

Dichos algoritmos se basan en el ajuste de las formas en las imágenes que contienen objetos a detectar. La problemática que representan estos métodos es la captura de la textura, además de requerir una correspondencia y convergencia local mínima ante las variaciones de la forma [14].

Este artículo investiga la detección de peatones combinando dos técnicas: la detección de movimiento por sustracción de fondo (DMSF) y modelos de forma activa (ASM), lo cual es una continuación natural de la investigación previa reportada por Antonio y Romero [15]. Contibuyendo al estado del arte, con un método novedoso para la detección de peatones, robusto ante variaciones de fondo, postura e indumentaria de las personas.

Por lo cual, se investiga la detección de peatones en escenas de video capturadas por una cámara fija, colocada en una variedad de interiores. El método aquí propuesto, primero detecta las regiones de movimiento y posteriormente, ajusta un modelo de forma activa en cada una de ellas. Nuestro método propuesto es robusto ante variaciones causadas al caminar, problemas de fondo, bajo contraste, y resolución de la imagen.

El proceso de detección inicia con el cálculo automático de los parámetros de traslación utilizando la técnica (*DMSF*), para localizar regiones de interés en los cuales *ASM* ajusta la silueta de un peatón. Para poder detectar y ajustar el contorno del peatón se utiliza una etapa que conlleva el uso de una pirámide de resolución [16], la cual disminuye la cantidad de iteraciones necesarias para ajustar el modelo de forma activa, debido a las variaciones de los niveles de grises, lo que hace que la combinación de las dos técnicas *DMSF* y *ASM*, hagan de esta investigación un tema poco explorado en el estado del arte [17-19].

I.1. TRABAJO RELACIONADO

La efectividad en la detección de movimiento depende de las variaciones en la videosecuencia, por ejemplo: ruido de origen, fondos complejos, variaciones en la iluminación de la escena y las sombras de los objetos estáticos y en movimiento. En el estado del arte se han sugerido varios métodos para superar estos problemas reteniendo solo el objeto móvil de interés. Éstos métodos han sido clasificados en tres categorías que son: sustracción de fondo, diferencia temporal y flujo óptico, [20–27]. Por otro lado, la sustracción de fondo proporciona una separación entre el fondo y los objetos de interés, pero es sensible a los cambios dinámicos debido a la iluminación y oclusión. El estado del arte categoriza la técnica de sustracción de fondo como: modelado básico de fondo, modelado estadístico de fondo, modelado de fondo difuso, agrupación de fondo, modelado de fondo por medio de una red neuronal,

modelado de fondo *wavelet* y estimación de fondo [28–33]. Por otro lado, Zheng et al., menciona que el método de flujo óptico puede ser usado para detectar objetos en movimiento. Sin embargo, la mayoría de los métodos de flujo óptico son computacionalmente complejos y no se pueden aplicar a transmisiones de video de *frames* completos en tiempo real sin tener un *hardware* especializado [34].

En Lee et al. [35], se propone un método que busca una región de fondo dinámica analizando el video obtenido desde una cámara *CCTV* y que ayuda a remover falsos positivos. Dicho método fue evaluado con el *dataset CDnet 2012/2014*, teniendo una precisión media de 86.50% y 76.68% en las bases de datos *CDnet2012* y *CDnet2014,* respectivamente.

Otro ejemplo de detección en sustracción de fondo ha sido investigado por Camplani y Salgado [36], quienes proponen una combinación de clasificadores que permiten mejorar la sustracción de fondo considerando que la imagen tenga factores de variabilidad como el color, sombras e iluminación. Su algoritmo ayuda a reducir falsas detecciones. Al final muestran un resultado con un valor medio de 70% utilizando la base de datos DcamSeq.

En Ramya y R. Rajeswar [37], presentan una nueva técnica que mejora el método de diferencia de frames al clasificar a nivel de píxeles los primeros frames como fondo y sus planos con un coeficiente de correlación, al final comparan sus resultados en diferentes imágenes originales, obteniendo el mejor resultado de precisión de 0.9984 en una imagen de camouflaje.

En una investigación basada en la detección de movimiento realizada por Sehairi et al. [38], quien presenta 12 técnicas basadas en detección de movimiento, en los cuales muestra los diferentes métodos de detección de movimiento como lo son: *temporal differencing* (*frame difference*), *three-frame difference* (*3FD*), *adaptive background* (average filter), *forgetting morphological temporal gradient* (*FMTG*), *background estimation*, *spatio-temporal markov field, running gaussian average* (*RGA*), *mixture of gaussians* (*MoG*), *spatio-temporal entropy image* (*STEI*), *difference-based spatio-temporal entropy image* (*DSTEI*), *eigen-background* (*Eig-Bg*) y *simplified self-organized map* (*Simp-SOBS*). Estos métodos son evaluados con el dataset *CDnet2014*, obteniendo el mejor resultado la técnica *GMM* con 0.99593 de especificidad, 3.08499 en *PWC* y 0.61021 de precisión. Mientras que la técnica de detección de movimiento con el menor rendimiento es *STEI*, con 0.78646 de especificidad, 22.18321 en *PWC* y 0.12881 de precisión.

Por otro lado, en el estado del arte basado en modelos de forma activa (*ASM*), se menciona una literatura amplia relacionada al ajuste de formas en órganos y partes del cuerpo humano, como manos o rostros [39–41], mas aún, en el trabajo realizado por Cootes et al. [42], mencionan la construcción de un modelo de distribución de puntos (*PDM*), donde se puede combinar la silueta del ser humano y sus articulaciones anatómicas de las coyunturas especificas y útiles para construir el modelo *ASM* y para poder segmentar posteriormente el cuerpo modelado en nuevas imágenes.

Por otro lado, la detección de peatones en videosecuencias utilizando modelos de forma activa en la literatura es limitada, debido a que se considera que requiere un alto consumo de tiempo de cómputo [16]; por lo cual la gran mayoría de las investigaciones realizadas se basan en la detección de personas estáticas (pose de pie sin moverse), y en pocos trabajos se menciona como se detectan personas caminando con *ASM*, lo que permite analizar la cadencia al caminar (*gait recognition*) [17].

En Baumberg y Hogg [43], se muestra como se puede dar seguimiento a un cuerpo no rígido en movimiento. Se dice que su detector funciona para poses y vistas aún con 2 o 3 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.

De igual forma Koschan et al. [44], aplican un modelo de forma activa 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 3 alineaciones de siluetas de una sola persona. Realizan detección de contorno y al final muestran el cálculo del error normalizado, entre los *landmarks* del *ground truth* y los *landmarks* estimados con un error medio en la distancia entre dichos puntos de 4.01 píxeles con 10 *landmarks*, 4.29 píxeles con 14 *landmarks*, 2.74 píxeles con 21 *landmarks* y 2.42 píxeles con 42 *landmarks* en la escena Man_6 y un error medio de 8.06 píxeles con 10 *landmarks*, 7.83 píxeles con 14 *landmarks*, 7.22 píxeles con 21 *landmarks* y 3.01 píxeles con 42 *landmarks* en la escena Man_9.

Por su parte en J. Y Jung [45], proponen un modelo con base a un exo-esqueleto para poder detectar poses humanas. Usan sustracción de fondo y un algoritmo de coincidencia de modelos de forma activa en el proceso de ajuste. Experimentan con 600 imágenes de personas con 17 *landmarks*, que muestran características de pose. La precisión reportada en tres diferentes poses son: 97.32%, 93.3% y 80.62%.

En Kim et al. [46], identifican la forma de caminar de los peatones, a partir de una secuencia de 122 *frames* de personas obtenidas de la base de datos *HGCD* [47]. Detectan peatones automáticamente con un modelo de forma utilizando 32 *landmarks* para la silueta de un humano a partir de imágenes de 720×480 *píxeles*. Al final obtienen una efectividad del 90% en la detección.

Lee y Choi [48], detectan peatones utilizando modelos de forma activa, usando imágenes en infrarrojo (*IR*) y cámaras visibles que permiten seguir peatones en entornos degradados y con poca luz. Marcan manualmente 42 *landmarks* alrededor del cuerpo pero no muestran resultados de sus experimentaciones.

Ma y Ren [49], proponen un método basado en modelos de forma activa para reconocer peatones. No especifican el número de *landmarks* para entrenar a su modelo. 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 de peatones.

Ide [50], propone una técnica para segmentar peatones empleando un modelo de forma activa entrenado con siluetas recortadas con la herramienta *grab-cut,* a la cual denominan:

ASFSeg, segmentación de retroalimentación de forma. Este método compara resultados de segmentación de *grab-cut* y los ajustes del modelo de forma activa eligiendo la mejor coincidencia como su segmentación. Reportan una tasa de error de 2.32% y 2.15% con *foreground* (*FG*) y *background* (*BG*), respectivamente.

Por su parte Vasconcelos y Tavares [51], detectan peatones que caminan en direcciones diversas. Para esto usan el conjunto de datos *CASIA Gait Database* [52], obtenidas de videos en formato *MPEG* con escenas de peatones en imágenes de 320×240 *píxeles*. Entrenan con 14 imágenes que representan la silueta del peatón y cada forma se representa por un total de 113 *landmarks*, 100 *landmarks* de la silueta, combinado con 13 *landmarks* pertenecientes a codos, rodillas y pies. Los resultados muestran un error de distribución medio de 4 a 7 *píxeles* en los 113 *landmarks* alrededor del contorno del peatón detectado.

II. PROCESO EXPERIMENTAL

Como se observa en la Figura 1, el proceso experimental de esta investigación consta de tres etapas: Detección de movimiento, ajuste del modelo de forma activa y evaluación.

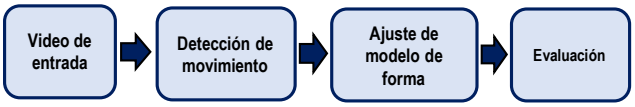


Fig. 1 Etapas del método propuesto: Detección de movimiento, ajuste del modelo de forma activa y evaluación.

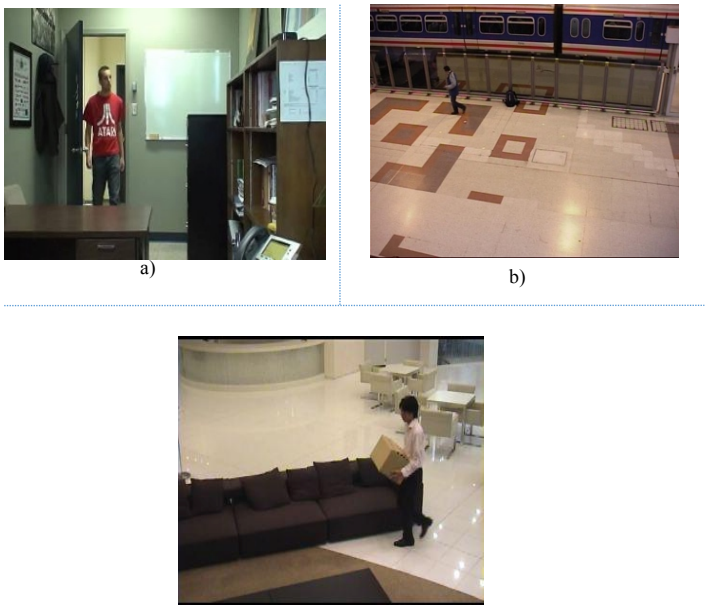
II.1 DATOS EXPERIMENTALES

En esta investigación se utilizan dos bases de datos del estado del arte *CDnet2014* y *CASIA Gait dataset*.

La base de datos *CDnet2014* contiene diversas escenas y ambientes, conteniendo entes en movimiento como peatones y autos. Las escenas de esta base de datos son útiles, puesto que contiene variaciones de iluminación, de fondo, y condiciones meteorológicas (e.g. lluvia, nieve, viento). En total, esta base de datos está compuesta por aproximadamente 90,000 *frames* contenidos en 31 videosecuencias. Seis categorías integran las videograbaciones de esta base de datos: *baseline*, *dynamic background*, *camera jitter*, *shadow*, *intermittent object motion* y *thermal*.

Por su parte, CASIA Gait dataset [53], es una base de datos que contiene un algoritmo de referencia y 12 experimentos, con lo cual investigan las formas de caminar de varios peatones. Sus imágenes contienen un entorno externo con fondo complejo, lo cual hace difícil el extraer las siluetas con calidad legible. Los doce experimentos fueron diseñados para evaluar la robustez de un algoritmo que detecte peatones con variación en la vestimenta. Los sujetos del video caminan en trayectoria elíptica y con un ángulo de visión que cambia conforme los sujetos van caminando, por lo que no se puede obtener la relación entre el ángulo de visión y el rendimiento del algoritmo. En la base de datos, se incluyen los datos adquiridos a partir de 11 vistas, consideran tres factores importantes: ángulo de visión, vestimenta y cambios en las condiciones al caminar.

Para esta investigación se seleccionan las escenas de *CDNet 2014* y de *CASIA Gait dataset.* En lo que se refiere a *CDNet 2014* se seleccionan tres videosecuencias que contienen peatones: *office*, *PETS2006* y *sofa* (ver Tabla 1). Las dos primeras videograbaciones pertenecen a la categoría de *baseline*, mientras que la tercera pertenece a la categoría *intermittent object motion* (ver Figura 2). Las tres videosecuencias seleccionadas contienen variaciones, principalmente de iluminación y de fondo. Las videograbaciones fueron capturadas en formato *MJPEG*, con una velocidad de 0.17 *frames* por segundo a una resolución de 360×240 *píxeles* en las escenas *office* y *sofa*, mientras que los *frames* de la escena *PETS2006* tienen una resolución de 720×576 *píxeles.*



a) b) c)

Fig. 2 Escenas de la base de datos *CDnet2014*: a) *office*, b) *PETS2006* y c) *sofa* utilizadas en esta investigación.

Tabla 1. Escenas seleccionadas y el número de frames utilizados para la construcción del conjunto de entrenamiento en el modelo de forma activa.

Escena	Número de *frames* a entrenar	*Frame* de inicio y final	Tamaño de *frame* en *píxeles*
office	50	610 - 660	360 × 240
PETS2006	50	1110 - 1159	720 × 576
sofa	50	105 - 154	360 × 240

En lo que se refiere a la bases de datos *CASIA Gait dataset,* ésta contiene videosecuencias donde se almacenan videos codificados con un tamaño de 320×240 *píxeles además* contiene información de 124 peatones los cuales caminan en diferentes direcciones. Para propósito de experimentación se seleccionaron 4 escenas (ver Figura 3), en las cuales caminan 4 sujetos que caminan en cuatro direcciones (0º, 36º, 54º y 90º).

II.1 DETECCIÓN DE MOVIMIENTO

Esta investigación parte con la premisa de que un peatón es

un ente en movimiento dentro de una escena. Por lo tanto, la primera etapa es detectar regiones de movimiento, para ello se propone la técnica de detección de movimiento con sustracción de fondo (*DMSF*), con la cual se separa el objeto de interés (*foreground*) del fondo (*background*) [34–36].

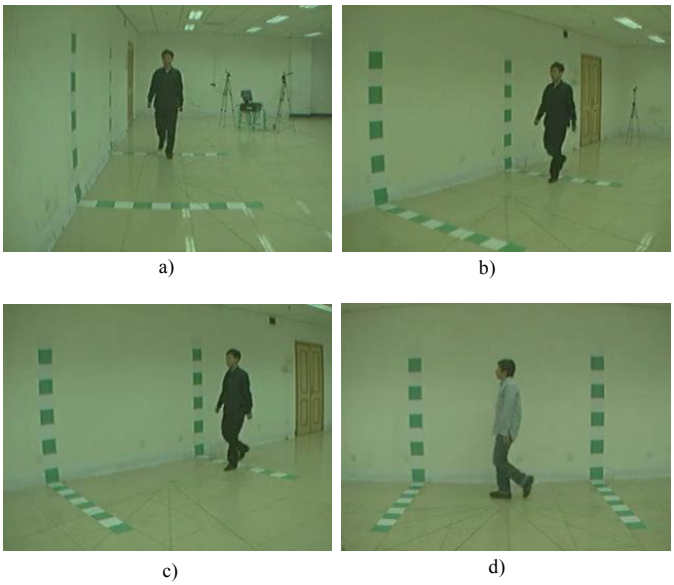


Fig. 3 Escenas de la base de datos *CASIA Gait dataset* utilizadas en esta investigación: a) *Peatón caminando a 0 grados*, b) *Peatón caminando a 36 grados*, c) *Peatón caminando a 54 grados y* d) *Peatón caminando a 90 grados*.

La Figura 4, ilustra nuestra implementación de la técnica *DMSF* para la detección de regiones de movimiento, la cual utiliza un *threshold* y operaciones morfológicas, con la finalidad de realizar una segmentación sin ruido y huecos.

Nuestra implementación del método de detección de movimiento con sustracción de fondo (DMSF), detecta regiones de cambio entre dos frames, cuyas etapas son las siguientes:

1. Convierte el primer frame de la videosecuencia a escala de grises, y es definido como frame de fondo (*F*).
2. Lee el siguiente frame en la videosecuencia (*I*) y se convierte a escala de grises.
3. Sustrae el fondo *F* de la imagen *I*, *B(x; y) = I(x, y) - F(x, y)*.
4. Binariza *B*, considerando un umbral *Th*, cuyo valor se calcula experimentalmente con base en el estado del arte, a través de la Ecuación 1 y una muestra de 50 imágenes de x×y píxeles.

$$Th = \frac{\sum_{i=1}^{x}\sum_{j=1}^{y} I(i,j)}{(x \times y)} \quad (1)$$

5. Erosiona *B*, utilizando un elemento estructurado circular E de radio *R = 5*, considerando que la silueta del cuerpo humano es curvilínea y una estructura circular ayuda a redondear el contorno.
6. Elimina residuos en *B* utilizando la técnica de 4 vecinos como se muestra en la Figura 5, entendiendo como residuo al conjunto de píxeles aislados de la silueta de la persona.
7. Rellena huecos en *B*, entendiendo que un hueco es un conjunto de píxeles de *B* con valor 0 al interior del contorno de la silueta de la persona.

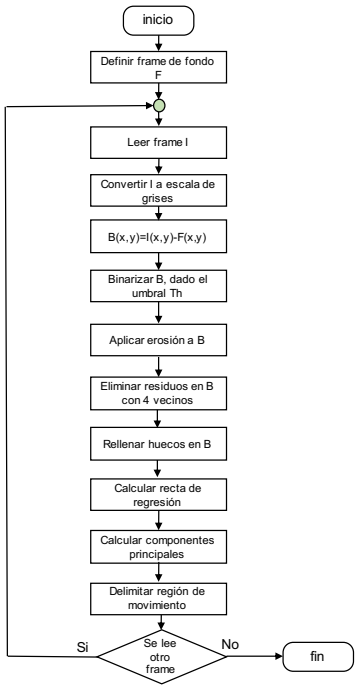


Fig. 4 Diagrama de flujo, donde se muestran las etapas para detectar una región de movimiento con la técnica *DMSF*.

Fig. 5. Eliminación de píxeles cercanos con la técnica de 4 vecinos.

8. Calcula la ecuación de la recta de regresión, la cual se usa en el paso 10 para delimitar el cluster donde se detectó movimiento (Ecuación 2):

$$y' = a_0 + a_1 x \quad (2)$$

Donde y_0 es la ordenada estimada de los *n* puntos *(x,y)*, que corresponden a los píxeles en *B* cuyo valor es 1, mientras que a_0 y a_1 son constantes que obtienen su valor de las ecuaciones normales de la recta de mínimos cuadrados, las cuales en esencia forman un sistema de dos ecuaciones con sus incógnitas cuya solución son las Ecuaciones 3 y 4.

$$a_0 = \frac{(\sum_{i=1}^{n} y_i)(\sum_{i=1}^{n} x^2) - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} x_i y_i)}{n(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)^2} \quad (3)$$

$$a_1 = \frac{n(\sum_{i=1}^{n} x_i y_i) - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)^2} \quad (4)$$

9. Calcula los componentes principales considerando *x* que representa los *n* puntos con coordenadas *(x,y)* que corresponden a los píxeles en *B* cuyo valor es 1.

El vector media μ está dado por Ecuación 5:

$$\mu = \frac{1}{n}\sum_{1=1}^{n} x_i \quad (5)$$

De manera similar, la matriz de covarianza n × n, Σ, se puede aproximar por Ecuación 6:

$$\Sigma = (x-\mu)(x-\mu)^T \quad (6)$$

Debido a que Σ es real y asimétrica, siempre es posible encontrar un conjunto de n eigenvectores ortonormales. Por lo tanto, la transformada de los componentes principales, también llamada transformada de Hotelling, se obtiene por Ecuación 7:

$$y = A(x-\mu) \quad (7)$$

Claramente, los elementos del vector y no están correlacionados. Por lo tanto, la matriz de *covarianzas* Σ es diagonal. Los renglones de A son los *eigenvectores* $[\vec{v}_1, \vec{v}_2]$ normalizados de Σ, donde los eigenvalores, λ_1 y λ_2, con $\lambda_1 \geq \lambda_2$. Debido a que A es real y simétrica, estos vectores forman una base *ortonormal*, y sus matrices inversa y transpuesta son iguales. Por lo tanto, se pueden obtener los valores de x a través de la transformación inversa Ecuación 8.

$$x = A^T(y+\mu) \quad (8)$$

10. Delimita la región de movimiento. Para esto, se calculan los puntos extremos: $P1(x_1, y_1)$, $P2(x_2, y_2)$, $P3(x_3, y_3)$ y $P1(x_4, y_4)$ de las rectas L_1 y L_2, cuyos vectores de dirección son los eigenvectores $\vec{v}_1$ y $\vec{v}_2$. Dichas rectas ortogonales se cruzan en el punto medio $O(\bar{x}_i, \bar{x}_i)$ de los píxeles en B cuyo valor es 1. Por consiguiente, como se muestra en la Figura 6, los puntos: $E1(x_4, y_1)$, $E2(x_2, y_1)$, $E3(x_2, y_3)$ y $E4(x_4, y_3)$ delimitan la región de movimiento. Para este caso, los puntos que forman la recta L1 también se pueden calcular utilizando la recta de regresión de la Ecuación 2.

II.2 Modelo de Forma Activa (*ASM*)

El modelo de forma activa (*ASM*), es un método flexible que ha sido usado para modelar y representar un amplio rango de objetos [56]. En la primera etapa de *ASM*, se realiza el modelo de distribución de puntos (*PDM*), éste especifica la forma media del objeto modelado además de las variaciones que éste pueda tener; para que la *PDM* sea robusta es necesario delimitar el contorno de la silueta mediante el marcado con los *landmarks* (ver Figura 7).

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*, con lo cual 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.

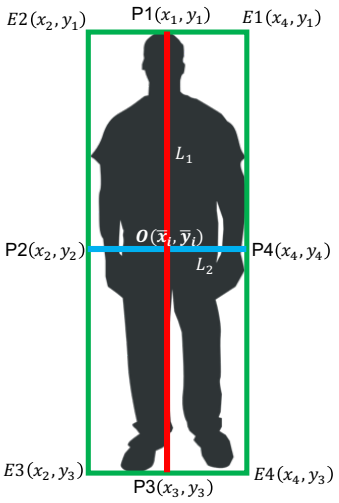


Fig. 6 Región de movimiento (rectángulo verde) delimitado por los puntos extremos: E1(x_4,y_1), E2(x_2,y_1), E3(x_2,y_3) y E4(x_4,y_3), de las rectas L_1 y L_2, cuyos vectores de dirección son los eigenvectores $\vec{v_1}$ y $\vec{v_2}$.

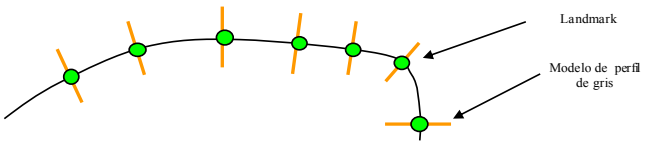


Fig. 7 Representación del modelado de puntos (landmarks, puntos verdes), y la obtención de los perfiles de grises (línea amarilla) de acuerdo con [57].

II.3 Ajuste del Modelo de Forma Activa

Una vez hallada la región de movimiento, donde se asume se encuentra un peatón, los puntos que delimitan esta región: el centroide y sus esquinas, se utilizan como referencia para ajustar un modelo de forma activa (ASM). Como se muestra en la Figura 8, el ajuste de un modelo de forma activa se realiza en dos fases, un entrenamiento donde se crea el modelo y una fase donde se ajusta este modelo.

Fase de entrenamiento

En esta fase se construye un modelo de distribución de puntos (PDM, por sus siglas en inglés), trabajo que se realiza en 5 etapas:

1. Obtener imágenes de entrenamiento. Para esta investigación se seleccionan escenas de las bases de datos *CDNet* 2014 y de *CASIA Gait dataset*. En lo que se refiere a *CDNet* 2014 se seleccionan tres videosecuencias con peatones: *office*, *PETS2006* y *sofa*; de las cuales las dos primeras pertenecen a la categoría de *baseline*, mientras que la tercera pertenece a la categoría *intermittent object motion*. Las tres videosecuencias seleccionadas contienen variaciones, principalmente de iluminación y de fondo. Las videograbaciones fueron capturadas en formato *MJPEG*, con una velocidad de 0.17 *frames* por segundo a una resolución de 360×240 *píxeles* en las escenas *office* y *sofa*, mientras que los *frames* de la escena *PETS2006*

tienen una resolución de 720×576 *píxeles*. En lo que se refiere a la base de datos *CASIA Gait dataset*, contiene videosecuencias donde se almacenan videos codificados con un tamaño de 320×240 *píxeles*; además, contiene información de 124 peatones los cuales caminan en 4 diferentes direcciones: 0º, 36º, 54º y 90º. Para la experimentación con esta base de dato se seleccionaron 4 escenas, en las cuales 4 sujetos caminan en las 4 direcciones.

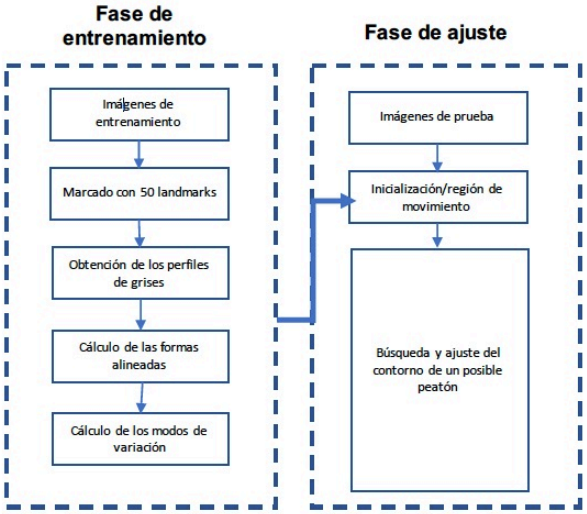


Fig. 8 Para el ajuste de un modelo de forma activa, se realizan en dos fases: entrenamiento y ajuste.

2. Marcado con 50 *landmarks*. Para esto se realiza el marcado de 50 puntos estratégicos alrededor del contorno de un peatón Figura 9, cosntruyendo un conjunto de entrenamiento donotado como:

$$x = [x_1, \dots, x_n, y_1, \dots, y_n]^T \quad (9)$$

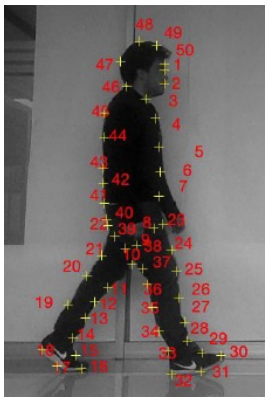


Fig. 9 Marcado con 50 ladmarks alrededor del contorno de un peatón.

3. Obtención de los perfiles de grises. Por cada uno de los 50 *landmarks* alrededor del contorno del modelo, se colectan los niveles de grises de los *píxeles ortogonales*. Como se muestra en la Figura 10, se obtiene un conjunto de puntos *ortonormales* en cada *landmark*, considerando 20 *píxeles* dentro y fuera de la silueta del peatón, respectivamente.

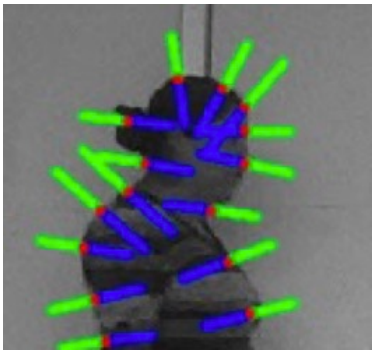


Fig. 10 Obtención de los perfiles de grises en cada uno de los 50 landmarks alrededor del peatón

4. Cálculo de las formas alineadas. La Figura 9, muestra el marcado de 50 *landmarks* en imágenes de entrenamiento y escala del peatón contenido en la imagen. Para comparar el par equivalente de coordenadas de las diferentes formas deben estar alineados con respecto a un sistema de referencia. Por lo tanto, se utilizan matrices de transformación geométrica, para aplicar escalado, rotación y traslación de las formas de entrenamiento, a fin de obtener formas alineadas Figura 11, minimizando el error $\boldsymbol{E}$ (Ecuación 10).

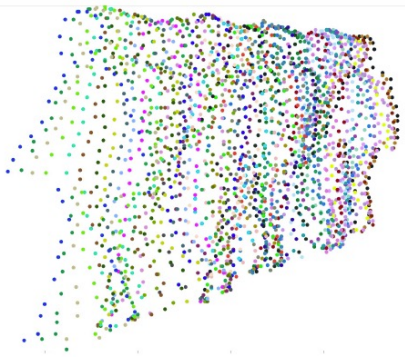


Fig. 11 Marcado de 50 landmarks en imágenes de entrenamiento

$$\mathrm{E} = (\mathrm{y} - \mathrm{Mx})^T \boldsymbol{W} (y - Mx) \quad (10)$$

donde **W** es una matriz diagonal cuyos elementos son factores de ponderación para cada punto de referencia y M representa la transformación geométrica de la rotación θ , la traslación **t** y escalado **s**. Los factores de pesos se establecen en relación con el desplazamiento entre las posiciones calculadas de los puntos de referencia antiguos y nuevos a lo largo del perfil. Si el desplazamiento es grande, entonces el factor de pesos correspondiente en la matriz se establece bajo; Si el desplazamiento es pequeño, entonces la ponderación con los pesos es alta. Dado un solo punto, denotado por $[x_0, y_0]^T$, la transformación geométrica se define como:

$$M = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = s \begin{bmatrix} cos\theta & sen\theta \\ -sen\theta & cos\theta \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \quad (11)$$

Después de aplicar la transformación geométrica, con los parámetros de pose, θ, **t**, **s**, se obtiene, la proyección de y en el modelo de coordenadas del *frame*, Ecuación 12.

$$x_p = M^{-1} y \quad (12)$$

Finalmente, los parámetros del modelo son actualizados como se define en la Ecuación 13:

$$b = \mathbf{\Phi}^T(x_p - \bar{x}) \quad (13)$$

Como resultado del procedimiento de búsqueda a lo largo de los perfiles, se obtiene el desplazamiento óptimo de un punto de referencia, con lo cual se genera una nueva forma en el marco de coordenadas de la imagen **y**, el conjunto de entrenamiento alineado se muestra en la Figura 12.

Fig. 12 Formas alineadas del conjunto de entrenamiento de un peatón

Cálculo de los modos de variación. Las coordenadas de los puntos (*landmarks*) del peatón para cada imagen de entrenamiento se almacenan y analizan estadísticamente para extraer las variaciones de forma, considerando que un conjunto de landmarks representa la forma del objeto. El conjunto de entrenamiento de esta investigación se compone de 50 formas diferentes, llamado conjunto de entrenamiento. Aunque cada forma en el conjunto de entrenamiento está en el espacio bidimensional, se puede modelar la forma con un número reducido de parámetros utilizando *análisis de componentes principales* (*PCA*). Considerando que se tienen m formas en el conjunto de entrenamiento x_I, para $i=1,..,m$. El *análisis de componentes principales* es el siguiente:

1. Se calcula la media de las m formas de la muestra en el conjunto de entrenamiento.
2. Se calcula la matriz de covarianza (**S**) del conjunto de entrenamiento.

$$S = \frac{1}{m}\sum_{i=1}^{m}(dx_i dx_i)^T \quad (15)$$

3. Se construye la matriz de *eigenvectores*.

$$\Phi = [\phi_1|\phi_2|\ldots|\phi_q], \quad (16)$$

Donde ϕ_j, $j=1,\ldots,q$ representa los *eigenvectores* de **S** correspondiente a los q más largos *eigenvalores*.

4. Dado que Φ y $\bar{x}$, donde cada forma pueda aproximarse como:

$$x_i \approx \bar{x} + \Phi b_i \quad (17)$$

donde:

$$b_i = \Phi^T(x_i - \bar{x}) \quad (18)$$

Para efectos de esta investigación se utilizan los q *eigenvalores* que representen el 98% de la variación de formas. Se obtienen así formas de peatón aceptables pudiéndose evaluar la distribución de ***b***, para restringir ***b*** a valores aceptables, donde se puede aplicar condiciones duras a cada elemento de $\boldsymbol{b_i}$ o con restricciones a ***b*** para ser representado en un *hiperelipsoide*.

Fase de ajuste

La fase de ajuste, busca adaptar el modelo de forma a una imagen de prueba, a través de las siguientes etapas:

1. Imágenes de prueba. Para llevar a cabo esta investigación dos tipos de conjuntos de datos fueron utilizados, *CDnet2014* y *CASIA Gait dataset* de los cuales se escogieron 50 imágenes para realizar la experimentación con validación cruzada.
2. Inicialización de la región de movimiento. En la etapa de ajuste del modelo ASM, es necesario obtener las variables de traslación (t_x, t_y), por lo cual se necesita obtener estas variables de traslación de la técnica *DMSF*, con lo cual se obtienen las coordenadas específicas para que *ASM* pueda realizar la búsqueda de los perfiles de grises y ajustar la silueta del peatón. Búsqueda y ajuste del contorno de un posible peatón. Con base a los parámetros de pose y de forma, se selecciona una imagen nueva (en escala de grises) que no pertenece al conjunto de entrenamiento; se busca hacer coincidir esta nueva forma en el conjunto de coordenadas de **x** (Ecuación 19). Para interpretar una forma dada en la imagen de entrada basada en el modelo de forma, se debe encontrar el conjunto de parámetros que mejor se adapte al modelo a la imagen. Suponiendo que el modelo de forma representa límites y bordes fuertes del objeto, un perfil en cada punto de referencia tiene una estructura local similar a un borde. Donde $g_{j,m}$; el cual es desplazado por ***m*** muestras a lo

$$\bar{x} = \frac{1}{m}\sum_{i=1}^{m} X_i \quad (19)$$

largo de la dirección normal del borde de la imagen correspondiente Ecuación 20.

$$f(\boldsymbol{g}_{j,m}) = (\boldsymbol{g}_{j,m} - \overline{\boldsymbol{g}}_j)^T \boldsymbol{S}_j^{-1}(\boldsymbol{g}_{j,m} - \overline{\boldsymbol{g}}_j) \quad (20)$$

Para obtener la pirámide de resolución del modelo de forma activa se realiza lo siguiente:

Cuando existen problemas con la longitud del perfil de gris, es necesario buscar el perfil de gris más cercano al *landmark*, donde la referencia del modelo debe estar cerca de su objetivo para poder ajustar el modelo. En

caso de que los perfiles de grises sean largos, la búsqueda para ajustar el modelo se vuelve computacionalmente costosa y hace que los niveles de grises se adhieran 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, haciendo que la imagen tenga desde una baja resolución a una alta resolución, generándolos con un suavizado gaussiano y submuestreo para producir una *pirámide de resolución* (Figura 13). El nivel 0 de la pirámide es la imagen original, 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 [58], de 1-5-8-5-1) y luego submuestreando cada *píxel* de gris correspondiente (ver Ecuación 21).

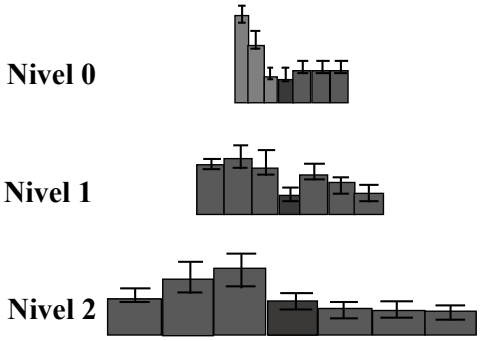


Fig. 13 Ejemplo de submuestreo en el cambio de resolución en una pirámide de resolución de una imagen [58].

$$y[n] = \sum_n x[2n]\, e^{-i\omega n} \qquad (21)$$

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 Ecuación 22.

$$b_g = P_g^T * y[n](g - \overline{g}) \qquad (22)$$

Donde $\mathbf{g}_{\text{ajustado}}$, representa **g** desplazado por $\boldsymbol{m}$ 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 (Figura 14).

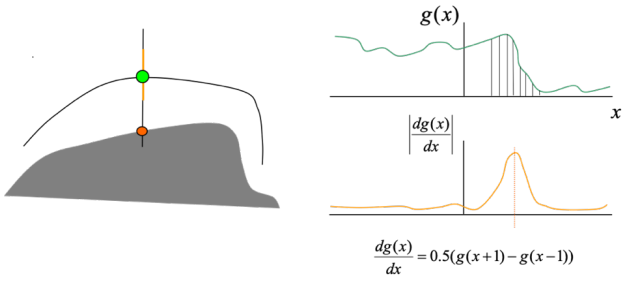


Fig. 14 Búsqueda de los perfiles de grises en un nuevo modelo aplicando el perfil de gris que coincida con el nuevo objeto.

III. Evaluación del modelo con las bases de datos CDnet 2014 y CASIA Gait Dataset

En esta sección se describe el proceso para la detección de peatones con variaciones de forma al caminar tomando en cuenta las bases de datos *CDNet 2014* y *CASIA GAIT DATASET*, donde se toman en cuenta dos fases: Entrenamiento (*PDM*) y ajuste (*ASM*). La fase *PDM* consiste en:

1. Alineación de las formas marcadas por los *landmarks* (Ec. 10-13).
2. Análisis de componentes principales (Ec. 14-18).
3. Construcción de niveles de grises, (Ec. 19).

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

1. Transformadas multiresolución aplicadas a las coordenadas de los contornos (Ec. 20).
2. Búsqueda activa y ajuste (Ec. 21).

Siguiendo este proceso, se construye un modelo de forma activa de un peatón.

En la etapa entrenamiento se usa el modelado de distribución de puntos (*PDM*), donde se propuso un marcado de 50 *landmarks* alrededor del contorno del peatón basándose de [59], tomando en cuenta algunos puntos antropométricos según el modelo *CAESAR* [60]. 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 *3D*. En las Figuras 15 y 16 se muestra a modo de ejemplo, las imágenes de los *frame 23* de un peatón entrando a la oficina (escena *office*), y el *frame 50* de un peatón caminando a cero grados, de la escena *CASIA Gait dataset*, cabe mencionar que cada imagen se les realizó un marcado manual (50 *landmarks*), en cada uno de los 50 *frames* formando un conjunto de entrenamiento, donde hay un peatón caminando en cada una de las escenas *office*, *PETS2006* y *sofa* de *CDNet 2014* y también de un peatón caminando a cero grados, peatón caminando a 34 grados, peatón caminando a 56 grados y peatón caminando a 90 grados de la escena *CASIA Gait dataset,* para ser usados en la etapa de *PDM*.

Posteriormente, se busca la alineación del conjunto de entrenamiento originando la forma media de todo el conjunto de entrenamiento. Para la captura de perfiles de grises se utilizó una recta perpendicular en cada *landmark* de 40 *píxeles* de ancho (20 *píxeles* hacia arriba y 20 *píxeles* hacia abajo), para utilizarlos en la búsqueda activa de grises en cada imagen del conjunto de entrenamiento. Se definió como criterio de parada, 2 *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.

El modelo de forma activa fue usado para construir un modelo que detecte a cada peatón en cada uno de los *frames* de las escenas tanto de *CDNet 2014* como de *CASIA Gait dataset* y que al momento de ajustar se definió como criterio de parada 2 iteraciones en la búsqueda de hasta 2 niveles de resolución en la búsqueda de los 20 *píxeles* en los perfiles de grises en cada *landmark* alrededor de la silueta de cada peatón.

En el proceso del modelado y para obtener el modelo de forma humana es necesaria la representación de los modos de variación más representativos de *PDM* en cada una de las

escenas (Figuras 17 y 18). Hay que mencionar que el primer modo de variación reúne la información sobre la posición de la forma de cada una de las personas, mientras que el segundo y tercer modo de variación obtienen la dirección en la que caminan además de que representan las diferentes deformaciones de las formas que se pueden adaptar en la búsqueda del ajuste.

Fig. 15 Ejemplo del marcado con 50 puntos característicos (*landmarks*), en la etapa de *PDM* de la escena *office* (*frame* 23) de CDNet 2014.

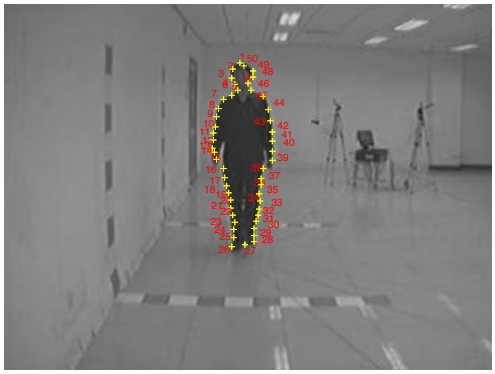


Fig. 16 Ejemplo del marcado con 50 puntos característicos (*landmarks*), en la etapa de *PDM* de la escena *peatón caminando a cero grados* (*frame* 50) de *CASIA Gait dataset*.

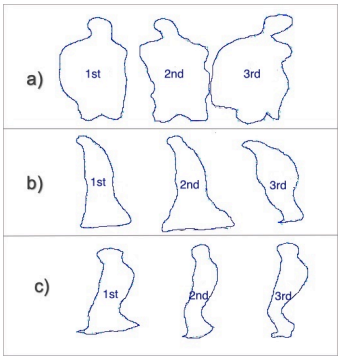


Fig. 17 Representación de 3 modos de variación del modelo de forma de los peatones de las escenas a) *office*, b) *PETS2006* y c) *sofa de CDNet2014*.

III.1 AJUSTE DEL MODELO DE FORMA ACTIVA CON LOS DATASET SELECCIONADOS

En esta etapa se busca detectar mediante el ajuste del modelo de forma de la silueta de un peatón en un *frame* en escala de grises, donde el algoritmo *ASM* busca la distancia mínima entre el *landmark* y los 20 *píxeles* de los perfiles de grises en cada uno de los 50 puntos alrededor del contorno del peatón.

Los resultados obtenidos de búsqueda del ajuste con *ASM* se muestran en los *frames* de las escenas *office*, *PETS2006* y *sofa* de *CDNet2014*, donde se muestran en las Figuras 19, 20 y 21, en los cuales se observa el marcado original (*GT*, *ground-truth*, puntos verdes) y el estimado, (puntos rojos).

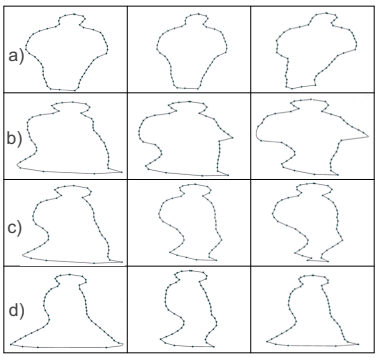
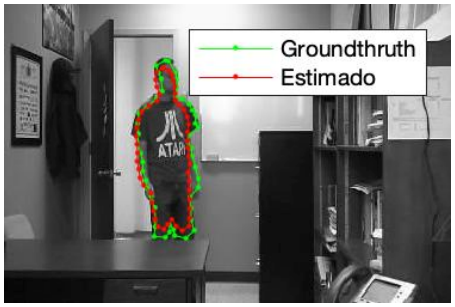
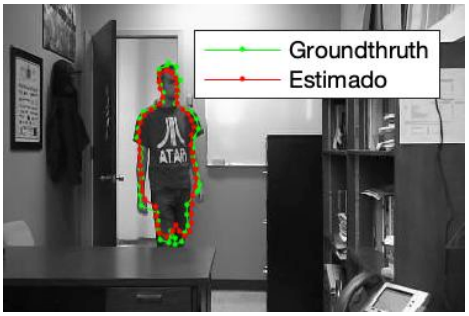


Fig. 18 Representación de los 4 modos de variación del modelo de forma de los peatones de las escenas a) *peatón caminando a cero grados* b) *peatón caminando a 36 grados*, c) *peatón caminando a 54 grados y d) peatón caminando a 90 grados de CASIA Gait dataset.*



a)

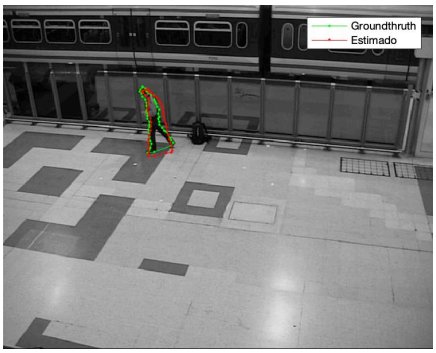


b)

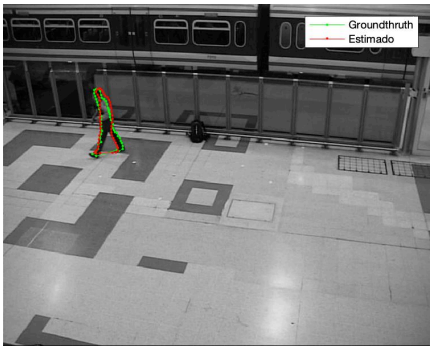
Fig. 19 Detección de un peatón en la escena *office* de CDnet2006 a) *frame* 630 y *frame* 634. Donde *GT* son los *landmarks* y los *landmarks* estimados.

En el análisis de la búsqueda del ajuste del peatón, los mejores ajustes se llevaron acabo en los *frames* donde no había contraste entre el peatón y el fondo, este contraste se presenta generalmente en la escena *sofa* donde el peatón el cual está vestido con un pantalón oscuro y se confundía con el sofá que también es de color oscuro. En la escena *PETS2006*, el mayor problema se presentó donde el peatón generaba un reflejo en el piso y el ajuste localizaba el reflejo como parte del peatón. Por su parte, en la escena *office* se presenta menos problemas de

ajuste, solamente donde existe oclusión entre el peatón y la puerta es donde el algoritmo no alcanza a realizar la detección.

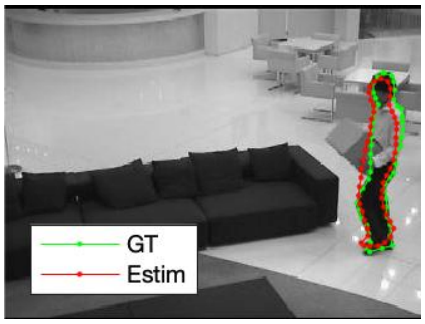


a)

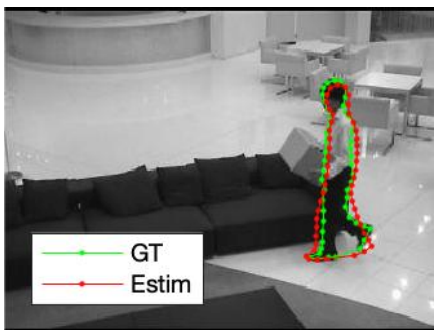


b)

Fig. 20 Detección de un peatón en la escena *PETS2006* de *CDnet2006* a) *frame* 1128 y *frame* 1151. Donde *GT* son los *landmarks del ground-truth* y los *landmarks* estimados.



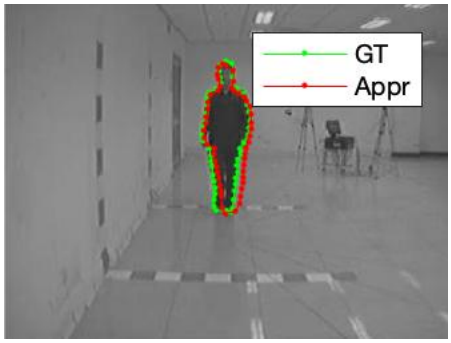
a)



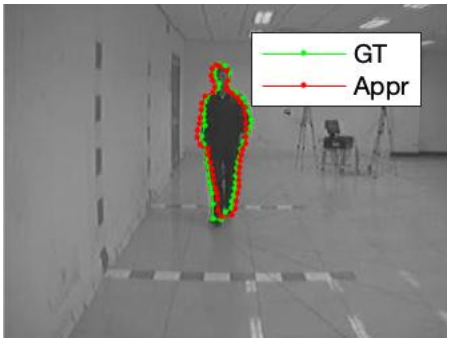
b)

Fig. 21 Detección de un peatón en la escena sofa de CDnet2006 a) *frame* 108 y *frame* 121 donde *GT* es el ground-truth y los *landmarks* estimados.

Los resultados obtenidos de búsqueda del ajuste con *ASM* se muestran en los *frames* de las escenas *peatón caminando a cero grados*, *peatón caminando a 36 grados*, *peatón caminando a 54 grados* y *peatón caminando a 90 grados* de *CASIA Gait dataset*. En la Figura 22, se muestran los resultados del ajuste de un peatón caminando a cero grados, como se puede ver en los resultados de ajuste como lo muestran los puntos rojos (aproximados) son muy cercanos a los *landmarks* del *ground truth* (puntos verdes).



a)



b)

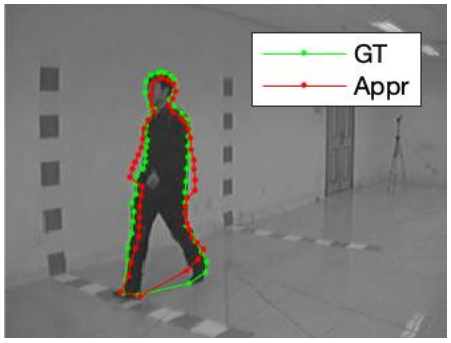
Fig. 22 Detección de un peatón caminando a cero grados de *CASIA Gait dataset* a) *frame* 38 y *frame* 43 donde *GT* es el ground-truth y los *landmarks* estimados.

En la Figura 23, se muestran los resultados del ajuste de un peatón caminando a 36 grados, como se observa el peatón camina a un ángulo en el cual hace difícil la grabación por la cámara y por ende la detección del peatón por el ángulo de la toma y por lo tanto los resultados de ajuste como lo muestran los puntos rojos (aproximados) tienden a estar lejanos en los brazos de los *landmarks* del *ground truth* (puntos verdes).

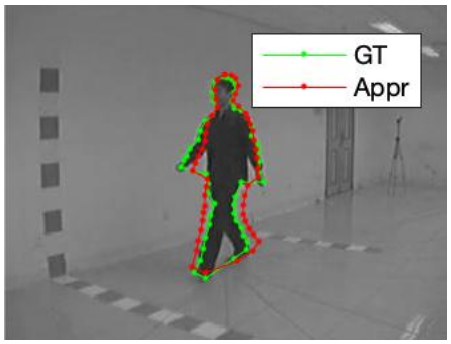
En la Figura 24, se muestran los resultados del ajuste de un peatón caminando a 54 grados, como se observa el peatón camina a un ángulo más pronunciado que en las escenas de un peatón caminando a 36 grados, lo cual hace difícil la grabación por la cámara y la detección del peatón por el ángulo de la toma, por lo tanto los resultados de ajuste como lo muestran los puntos rojos (aproximados) tienden a estar lejanos en las piernas de los *landmarks* del *ground truth* (puntos verdes).

En la Figura 25, se muestran los resultados del ajuste de un peatón caminando a 90 grados, como se observa el peatón caminando de perfil muy diferentes a las escenas de un peatón caminando a 36 grados o a 54 grados, lo cual hace difícil la grabación por la cámara y la detección del peatón por el ángulo de la toma, por lo tanto los resultados de ajuste como lo

muestran los puntos rojos (aproximados) tienden a estar lejanos en las piernas de los *landmarks* del *ground truth* (puntos verdes).

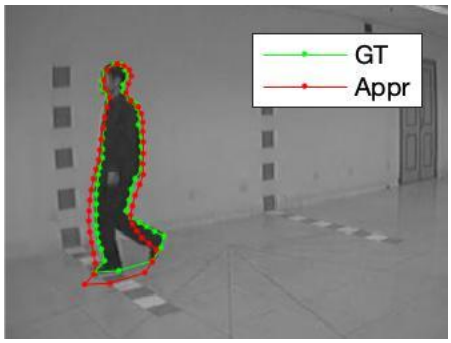


a)

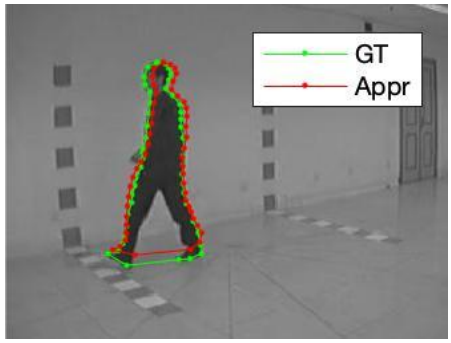


b)

Fig. 23 Detección de un peatón caminando a 36 grados de *CASIA Gait dataset* a) *frame* 24 y *frame* 27 donde *GT* es el ground-truth y los *landmarks* estimados.

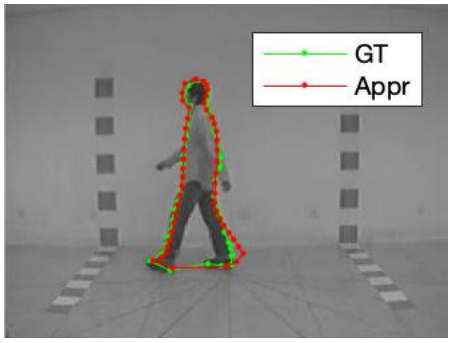


a)

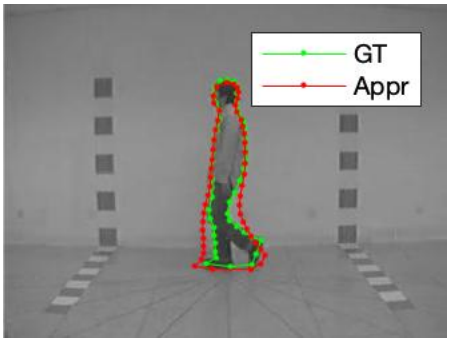


b

Fig. 24 Detección de un peatón caminando a 54 grados de CASIA Gait dataset a) *frame* 36 y *frame* 38 donde *GT* es el ground-truth y los *landmarks* estimados.



a)



b)

Fig. 25 Detección de un peatón caminando a 90 grados de CASIA Gait dataset a) *frame* 24 y *frame* 27 donde *GT* es el ground-truth y los *landmarks* estimados.

IV. RESULTADOS

Para verificar la precisión del ajuste logrado por el modelo de forma activa, se realiza una validación cruzada *leave one out* [61], para calcular el error de ajuste se utiliza como métrica la distancia euclidiana (e) entre los puntos (x_g,y_g) del *ground truth* y los puntos ajustados (x_a,y_a) obtenidos por el ajuste del modelo de forma activa.

Entonces, el cálculo del error de ajuste por cada punto del modelo de forma activa está dado por :

$$e = \sqrt{(x_g - x_a)^2 + (y_g - y_a)^2} \tag{21}$$

Por tanto, el cálculo del error de ajuste medio es:

$$\bar{e} = \frac{1}{K}\sum_{j=1}^{K} e_k \tag{22}$$

Donde $k=50$, que corresponde al número de iteraciones de la validación cruzada.

Finalmente, se reporta la gran media de los errores de ajuste de las 50 iteraciones de la validación cruzada a través de diagramas de caja. Considerando que un diagrama de caja es una herramienta útil para analizar el nivel de dispersión de los errores de ajuste obtenidos. Se presenta un diagrama de caja por cada uno de los conjuntos de imágenes experimentales de la base de datos *CDnet2014* donde cada *frame* tiene una resolución de *360×240*, en las escenas de *office* y *sofa;* de *720×576 píxeles* en la escena *PETS2014* y *CASIA Gait dataset*,

donde todos los *frames* tienen una resolución de *320×240 píxeles*.

En la evaluación de las escenas de la base de datos *CDNet2014*, en la escena *office*, se obtuvo un error de ajuste medio de 7.2 *píxeles*, el cual es un resultado satisfactorio considerando el nivel de oclusión en las extremidades inferiores del peatón en la escena.

En lo que se refiere a la escena *PETS2006*, a pesar de las dificultades que tiene con el brillo del piso, las sombras y oclusión, se obtiene un error de ajuste medio de 8.3 *pixeles*. De forma similar, en la escena *sofa* se obtuvo un error de ajuste medio de 7.1 píxeles.

Finalmente, la Tabla 2 y Figura 26, resumen los errores de ajuste medio obtenidos en la experimentación de las escenas *office*, *PETS2006* y *sofa* de *CDnet2014*.

Tabla 2 Resumen del promedio total del cálculo del error de ajuste en las escenas *office*, *PETS2006* y *sofa*. Además del número de iteraciones y la cantidad de pixeles en los perfiles de grises obtenidos en cada *landmark*.

Escena	Total de *frames*	Numero de perfiles de grises	Error de ajuste medio	Número de iteraciones
office	50	40	7.2	2
PETS2006	50	40	8.3	2
sofa	50	40	7.1	2

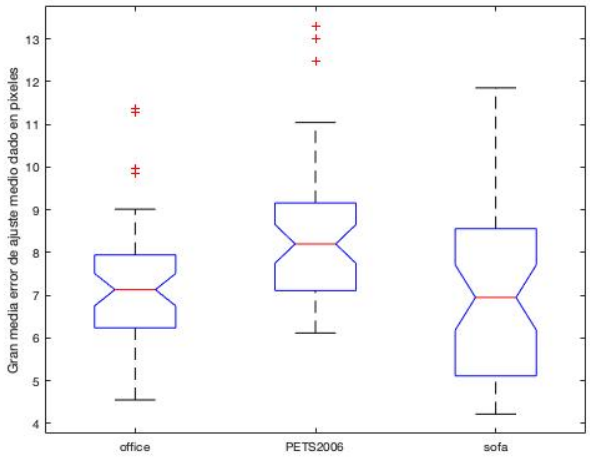


Fig. 26 Representación del error de ajuste medio entre los *landmarks* del *ground-truth* y los *landmarks* estimados en las escenas *office*, *PETS2006* y *sofa* de *CDnet2014*.

En la evaluación de las escenas de la base de datos *CASIA Gait dataset*, en la escena *peatón caminando a cero grados*, con el cual se obtuvo un error de ajuste medio de 4.5 *píxeles*, el cual es un resultado satisfactorio considerando que esta escena no contiene variaciones relevantes al caminar.

En lo que se refiere a la escena *peatón caminando a 36 grados*, a pesar de las dificultades que tiene con el ángulo, se obtiene un error de ajuste medio de 6.4 pixeles. De forma similar, en la escena *peatón caminando a 54 grados* se obtuvo un error de ajuste medio de 6.1 píxeles. Por último en la escena *peatón caminando a 90 grados,* se obtuvo un resultado de error de ajuste medio de 5.6 píxeles, debido a que en esta escena se observan variaciones al abrir el compás de las piernas.

Finalmente, la Tabla 2 y Figura 27, resumen los errores de ajuste medio obtenidos en la experimentación de las escenas experimentadas de la base de datos *CASIA Gait dataset*.

Escena	Total de *frames*	Numero de perfiles de grises	Error de ajuste medio	Número de iteraciones
peatón 0 grados	50	40	4.5 píxeles	2
peatón 36 grados	50	40	6.4 pixeles	2
peatón 54 grados	50	40	6.1 píxeles	2
peatón 90 grados	50	40	5.6 píxeles	2

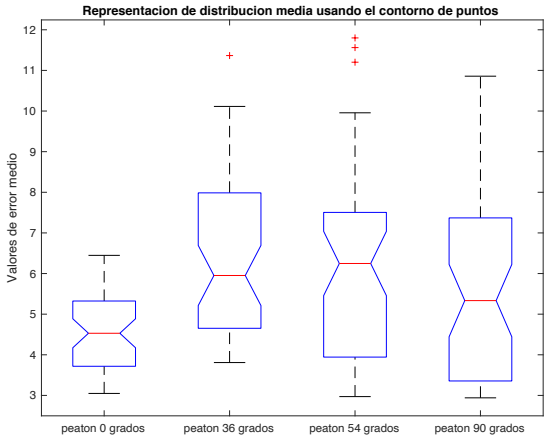


Fig. 27 Representación del error de ajuste medio entre los *landmarks* del *ground-truth* y los *landmarks* estimados en las escenas *peatón caminando a 0 grados*, *peatón caminando a 36 grados, peatón caminando a 54 grados* y *peatón caminando a 90 grados* de la base de datos *CASIA Gait dataset*.

V. DISCUSIÓN Y CONCLUSIONES

En este trabajo se propuso la ubicación de peatones con problemas de fondo, al detectar el movimiento con dos técnicas la técnica de detección de movimiento y el modelo de forma activa (*ASM*). Se evaluaron 2 bases de datos *CDnet2014 y CASIA Gait dataset,* de estas bases de datos se seleccionaron los *frames* que se utilizaron para entrenar a cada conjunto de 50 imágenes mediante el modelado de distribución de puntos (*PDM*), con 50 *landmarks* alrededor del contorno del cuerpo de cada peatón en las escenas mencionadas. Al final se muestran los mejores resultados dando un promedio total en el cálculo del error de ajuste medio, 7.5 *píxeles* en la búsqueda del ajuste de cada peatón de las escenas de la base de datos CDNet2014 y un resultado en el cálculo del error de ajuste medio de 5.6 píxeles en las escenas de la base de datos de *CASIA Gait dataset.*

La presente investigación se comparó con el trabajo realizado por Vasconcelos et al., [51], donde utilizan una base de datos tomada de *CASIA Gait dataset*, donde cada uno de los *frames* tienen una resolución de *340×240 píxeles*, con una tasa de 25 *fps*. Las imágenes contienen peatones caminando en diferentes direcciones, pero el fondo de cada uno de los *frames* es claro y los peatones tienen un alto contraste por lo cual la detección con un método de detección de movimiento además del algoritmo *ASM* hacen más fácil el ajuste de la silueta de cada peatón. Otros puntos de discusión del trabajo de Vasconselos son: (a) utiliza 2743 *frames* de entrenamiento, (b) construye un modelo de 113 *landmarks*, (c) utiliza un criterio de parada de 16 iteraciones; para poder ajustar un peatón y así obtener como resultado un error medio de 4 a 5 *píxeles*.

A diferencia del trabajo mencionado, en esta investigación se entrenó con una cantidad menor de *landmarks* (50 *landmarks*), con un criterio de parada de 2 iteraciones, además de utilizar una base de datos con características diferentes a la de *CASIA Gait dataset*. *CDnet2014* tiene imágenes que incluyen problemas de fondo, escala, resolución, además de otras características que hacen díficil la detección.

En trabajos futuros se espera detectar peatones con mayores variaciones al caminar, pero esta vez utilizando un conjunto de datos mayor y diferentes al utilizados en esta investigación donde contenga escenas con diferentes problemas de oclusión, iluminación y problemas de fondo.

AGRADECIMIENTOS

Agradecemos a CONACYT por la asignación de la beca con número CVU/Becario 289380/623403 para estudios de posgrados.

REFERENCIAS

[1] A. Jordão and W. R. Schwartz, "The Good, The Fast and The Better Pedestrian Detector," *Univ. Fed. Minas Gerais - Dep. Ciência da Comput.*, vol. 1, pp. 1–51, 2016.

[2] A. T. Angonese and P. F. Ferreira Rosa, "Multiple people detection and identification system integrated with a dynamic simultaneous localization and mapping system for an autonomous mobile robotic platform," *ICMT 2017 - 6th Int. Conf. Mil. Technol.*, pp. 779–786, 2017.

[3] D. S. Kim and K. H. Lee, "Infrared Physics & Technology Segment-based region of interest generation for pedestrian detection in far-infrared images," *INFRARED Phys. Technol.*, vol. 61, pp. 120–128, 2013.

[4] A. Halidou, X. You, M. Hamidine, R. A. Etoundi, L. H. Diakite, and Souleimanou, "Fast pedestrian detection based on region of interest and multi-block local binary pattern descriptors," *Comput. Electr. Eng.*, 2014.

[5] A. Lakshmi, A. G. J. Faheema, and D. Deodhare, "Pedestrian detection in thermal images: An automated scale based region extraction with curvelet space validation," *Infrared Phys. Technol.*, 2016.

[6] P. Dollár, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2012.

[7] A. Hill, A. Thornham, and C. J. Taylor, "Model-Based Interpretation of 3D Medical Images," 2013.

[8] A. Blake, R. Curwen, and A. Zisserman, "A framework for spatiotemporal control in the tracking of visual contours," *Int. J. Comput. Vis.*, 1993.

[9] C. G. Keller, M. Enzweiler, and D. M. Gavrila, "A new benchmark for stereo-based pedestrian detection," in *IEEE Intelligent Vehicles Symposium, Proceedings*, 2011.

[10] C. Hilario, J. M. Collado, J. M. Armingol, and A. De La Escalera, "Pedestrian detection for intelligent vehicles based on active contour models and stereo vision," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2005.

[11] A. Pentland and B. Horowitz, "Recovery of Non-Rigid Motion and Structure," *IEEE Trans. Pattern Anal. Mach. Intell.*, 1991.

[12] N. Vandenbroucke, L. Macaire, C. Vieren, and J. G. Postaire, "Contribution of a color classification to soccer players tracking with snakes," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, 1997.

[13] S. Zhang, C. Bauckhage, and A. B. Cremers, "Efficient Pedestrian Detection via Rectangular Features Based on a Statistical Shape Model," *IEEE Trans. Intell. Transp. Syst.*, 2015.

[14] F. Flohr and D. Gavrila, "PedCut: an iterative framework for pedestrian segmentation combining shape models and multiple data cues," in *Procedings of the British Machine Vision Conference 2013*, 2013.

[15] M. Antonio Juan Alberto y Romero, "Detección de peatones con variaciones de forma al caminar con Modelos de Forma Activa," *Cienc. ergo-sum*, vol. 27, no. 3, p. 17, 2020.

[16] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models - their training and application," *Comput. Vis. Image Underst.*, 1995.

[17] S. Das Choudhury and T. Tjahjadi, "Gait recognition based on shape and motion analysis of silhouette contours," *Comput. Vis. Image Underst.*, 2013.

[18] J. Müller and M. Arens, "Human pose estimation with Implicit Shape Models," in *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*, 2010.

[19] K. Ogawara, X. Li, and K. Ikeuchi, "Marker-less human motion estimation using articulated deformable model," in *Proceedings - IEEE International Conference on Robotics and Automation*, 2007.

[20] M. Piccardi, "Background subtraction techniques: A review," in *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, 2004.

[21] I. S. Kim, H. S. Choi, K. M. Yi, J. Y. Choi, and S. G. Kong, "Intelligent visual surveillance - A survey," *Int. J. Control. Autom. Syst.*, 2010.

[22] M. Paul, S. M. E. Haque, and S. Chakraborty, "Human detection in surveillance videos and its applications - a review," *Eurasip Journal on Advances in Signal Processing*. 2013.

[23] D. A. Migliore, M. Matteucci, and M. Naccari, "A revaluation of frame difference in fast and robust motion detection," in *Proceedings of the ACM International Multimedia Conference and Exhibition*, 2006.

[24] S. Yalamanchili, W. N. Martin, and J. K. Aggarwal, "Extraction of moving object descriptions via differencing," *Comput. Graph. Image Process.*, 1982.

[25] R. Jain, W. N. Martin, and J. K. Aggarwal, "Segmentation through the detection of changes due to motion," *Comput. Graph. Image Process.*, 1979.

[26] B. D. Lucas and T. Kanade, "ITERATIVE IMAGE REGISTRATION TECHNIQUE WITH AN APPLICATION TO STEREO VISION.," 1981.

[27] B. K. P. Horn and B. G. Schunck, "'Determining optical flow': a retrospective," *Artif. Intell.*, 1993.

[28] T. Bouwmans, "Traditional and recent approaches in background modeling for foreground detection: An overview," *Computer Science Review*. 2014.

[29] W. X. Kang, W. Z. Lai, and X. B. Meng, "An adaptive background reconstruction algorithm based on inertial filtering," *Optoelectron. Lett.*, 2009.

[30] S. Jiang and Y. Zhao, "Background extraction algorithm base on Partition Weighed Histogram," in *Proceedings - 2012 3rd IEEE International Conference on Network Infrastructure and Digital Content, IC-NIDC 2012*, 2012.

[31] L. Maddalena and A. Petrosino, "A self-organizing approach to background subtraction for visual surveillance applications," *IEEE Trans. Image Process.*, 2008.

[32] B. Antić, V. Crnojević, and D. Ćulibrk, "Efficient wavelet based detection of moving objects," in *DSP 2009: 16th International Conference on Digital Signal Processing, Proceedings*, 2009.

[33] S. Messelodi, C. M. Modena, N. Segata, and M. Zanin, "A Kalman filter based background updating algorithm robust to sharp illumination changes," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2005.

[34] J. Zheng, Y. Wang, N. L. Nihan, and M. E. Hallenbeck, "Extracting roadway background image: Mode-based approach," in *Transportation Research Record*, 2006.

[35] S. H. Lee, G. C. Lee, J. Yoo, and S. Kwon, "WisenetMD: Motion

detection using dynamic background region analysis," *Symmetry (Basel).*, 2019.
[36] M. Camplani and L. Salgado, "Background foreground segmentation with RGB-D Kinect data: An efficient combination of classifiers," *J. Vis. Commun. Image Represent.*, 2014.
[37] P. Ramya and R. Rajeswari, "A Modified Frame Difference Method Using Correlation Coefficient for Background Subtraction," in *Procedia Computer Science*, 2016.
[38] K. Sehairi, C. Fatima, and J. Meunier, "A Benchmark of Motion Detection Algorithms for Static Camera: Application on CDnet 2012 Dataset," in *Lecture Notes in Networks and Systems*, 2019.
[39] T. Huysmans, P. Moens, and R. Van Audekercke, "An active shape model for the reconstruction of scoliotic deformities from back shape data," *Clin. Biomech.*, 2005.
[40] N. Razali and A. Wahab, "2D Affective Space Model (ASM) for detecting autistic children," in *Proceedings of the International Symposium on Consumer Electronics, ISCE*, 2011.
[41] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T. S. Huang, "Interactive facial feature localization," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2012.
[42] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Training Models of Shape from Sets of Examples," in *BMVC92*, 1992.
[43] A. M. Baumberg and D. C. Hogg, "An efficient method for contour tracking using active shape models," *Proc. 1994 IEEE Work. Motion Non-rigid Articul. Objects*, 1994.
[44] A. Koschan, S. Kang, J. Paik, B. Abidi, and M. Abidi, "Color active shape models for tracking non-rigid objects," *Pattern Recognit. Lett.*, 2003.
[45] C. J. and K. Jung, "Human Pose Estimation ASM." World Academy of Science, Engineering and Technology, pp. 530–534, 2008.
[46] D. Kim, S. Lee, and J. Paik, "Active shape model-based gait recognition using infrared images," *Commun. Comput. Inf. Sci.*, vol. 61, no. 4, pp. 275–281, 2009.
[47] H. Sadoghi Yazdi, H. J. Fariman, and J. Roohi, "Gait Recognition Based on Invariant Leg Classification Using a Neuro-Fuzzy Algorithm as the Fusion Method," *ISRN Artif. Intell.*, 2012.
[48] D. Lee and S. Choi, "Multisensor fusion-Based object detection and tracking using Active Shape Model," *2011 6th Int. Conf. Digit. Inf. Manag. ICDIM 2011*, pp. 108–114, 2011.
[49] J. Ma and F. Ren, "Detect and track the dynamic deformation human body with the active shape model modified by motion vectors," in *2011 IEEE International Conference on Cloud Computing and Intelligence Systems*, 2011, pp. 587–591.
[50] I. Ide, "Segmentation of Human Instances Using Grab-cut and Active Shape Model Feedback," pp. 11–14, 2013.
[51] M. J. M. Vasconcelos and J. M. R. S. Tavares, "Human motion segmentation using active shape models," *Lect. Notes Comput. Vis. Biomech.*, 2015.
[52] K. Arai and R. Andrie, "Gait recognition method based on wavelet transformation and its evaluation with Chinese Academy of Sciences (CASIA) gait database as a human gait recognition dataset," in *Proceedings of the 9th International Conference on Information Technology, ITNG 2012*, 2012.
[53] S. Yu, D. Tan, and T. Tan, "A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition," in *Proceedings - International Conference on Pattern Recognition*, 2006.
[54] Z. Xu, B. Min, and R. C. C. Cheung, "A robust background initialization algorithm with superpixel motion detection," *Signal Process. Image Commun.*, 2019.
[55] J. Yin, L. Liu, H. Li, and Q. Liu, "The infrared moving object detection and security detection related algorithms based on W4 and frame difference," *Infrared Phys. Technol.*, 2016.
[56] T. F. Cootes and C. J. Taylor, "Active Shape Models — 'Smart Snakes,'" in *BMVC92*, 1992.
[57] I. M. Scott, T. F. Cootes, and C. J. Taylor, "Improving appearance model matching using local image structure," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 2003.
[58] T. F. Cootes, C. J. Taylor, and A. Lanitis, "Active Shape Models: Evaluation of a Multi-Resolution Method for Improving Image Search," 2013.
[59] A. Godil, "Advanced human body and head shape representation and analysis," *Digit. Hum. Model. - Springer Berlin Heidelb.*, pp. 92–100, 2007.
[60] S. Ressler, "A Web-based 3D Glossary for Anthropometric Landmarks," *Proc. HCI Int.*, vol. 1, pp. 1–5, 2001.
[61] A. Vehtari, A. Gelman, and J. Gabry, "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC," *Stat. Comput.*, 2017.

Juan Alberto Antonio. He was born on January 23, 1975, in Mexico City, Mexico. He received an Isthmus Technological Institute, Juchitan Oaxaca in 2000th, from a computer systems engineer, master in Center for Innovation and Technological Development in Computing, IPN in 2010th; and he is a PhD student in engineering sciences. His research topic is based on pedestrian detection and motion detection techniques**.**

Marcelo Romero, is researcher professor in computer science at the Autonomous University of the State of Mexico. He obtained his PhD from the University of York in 2010, United Kingdom. His research topics are anthropometric points, image processing, object detection, pattern recognition, computing in mathematics, natural Science, engineering and medicine, artificial neural network and theory of computation, etc.

Roberto Alejo, is researcher professor in machine learning and computer science, at the Technological Institute of Toluca. He studied Computer Systems Engineer from the Technological Institute of Toluca, Master of Science, in Computer Science from the Technological Institute of Toluca. PhD in Advanced Informatic Systems from the Universitat Jaume I, Spain. His line of research is based on the application of data mining techniques, machine learning, image processing, classification, pattern recognition, computer vision, feature extraction and algorithms.

CAPÍTULO 5

Conclusiones y trabajo futuro

Este capítulo concluye esta tesis y presenta posibles líneas de investigación como trabajo futuro.

5.1. Conclusiones

En esta tesis se investigó la detección de peatones desde escenas de videovigilancia. Específicamente, la propuesta realizada combina el potencial de dos técnicas del estado de arte: la detección de movimiento y los modelos de forma activa. La técnica de detección de movimiento utilizada se basa en la diferencia de frames con sustracción de fondo (*DFSF*). Esta técnica es utilizada en los sistemas de videovigilancia, debido a que se presenta robusta ante problemas de fondo, iluminación, sombras, etc. Al aplicar esta técnica en cada frame de una videosecuencia, se identifican las regiones que hayan tenido algún cambio, las cuales se etiquetan y sirven de entrada para ajustar un modelo de forma activa de un peatón. El modelo de forma activa (*ASM*) fue ejecutado para evaluar las variaciones de forma desde una escena de video donde aparecieron peatones videograbados. Para esto fueron aplicados los pasos en el modelo del algoritmo *ASM* y sus etapas fueron: entrenamiento *PDM*, extracción de la escala de grises y la búsqueda activa en una imagen nueva que no pertenece al conjunto de entrenamiento. En la etapa de *PDM*, se realizó un marcado a cada peatón en los diferentes frames, la cantidad de *landmarks* que se utilizaron para anotar fueron 50 alrededor de la silueta de un peatón tomando en cuenta los puntos antropométricos esenciales. En la última etapa del algoritmo *ASM* se buscó el ajuste de un peatón a pesar de las variaciones al caminar. Al final, *ASM* permitió la localización de peatones en las escenas, para lo cual se requirió un conjunto de entrenamiento estadísticamente significativo. Durante la experimentación, se observó que el ajuste del modelo de forma activa requiere un tiempo de procesamiento considerable, además de que disminuye su precisión ante variaciones de fondo extremas. En esta investigación se utilizó la base de

datos *CDnet2014*, la cual esta basada en escenas de videovigilancia. Este *dataset* contiene escenas de peatones, autos y otros objetos en movimiento; todos ellos contenidos en escenas con variaciones de fondo, iluminación y ruido. Tres secuencias de la base de datos *CDnet2014* fueron experimentadas: *Office*, *PETS2006* y *sofa*; las cuales presentan tanto variaciones en sus escenas, así como diferente número de peatones.
Por lo antes mencionado, se cumple el objetivo general planteado para esta tesis, aceptándose la hipótesis postulada dado que el error de ajuste medio obtenido en nuestra experimentación es menor a 10 píxeles.

5.2. Trabajo futuro

La propuesta de esta tesis basada en la detección de movimiento por sustracción de fondo y el ajuste de un modelo de forma activa de un peatón tuvo resultados satisfactorios, sin embargo, existen aún oportunidades de mejora, por lo que se proponen las siguientes líneas de investigación:

1. Mejorar la detección de las regiones de movimiento en cada *frame* de una videosecuencia, experimentando con técnicas alternas a la sustracción de fondo. Esto es recomendable puesto que es limitada la aplicación de un solo *frame* como fondo para una escena, sobre todo si esa escena es una grabación que cubre periodos de tiempo largos, como días o semanas. En esos casos, las variaciones en objetos, iluminación y condiciones climáticas (en grabaciones de exteriores), afectadas de forma mediata las escenas. En una investigación preliminar, se pudo verificar la viabilidad de las técnicas: Diferencia temporal (*Temporal differencing*), diferencia en tres *frames* (*Three-frame differencing*), y sustracción de fondo adaptado (*Adaptive background substraction*).

2. Utilizar algoritmos de detección de objetos alternos a los modelos de forma activa. En esta investigación se ha observado que es posible lograr un rendimiento aceptable en la localización de peatones utilizando esta técnica, sin embargo, las variaciones extremas en el fondo y la indumentaria de los transeúntes generan cambios para los cuales el modelo de forma no fue entrenado, lo que incide en un ajuste deficiente. Actualmente, en el estado del arte se esta investigando la viabilidad de las redes neuronales convolucionales en diferentes problemas de detección de objetos, por lo tanto, éstas representan una opción viable para investigar.