Evaluación de desempeño del descriptor de covarianzas en la detección de objetivos

Pedro Cortez Cargill*, Cristobal Undurraga Rius*, Domingo Mery Quiroz* y Alvaro Soto*
*Departamento de Ciencias de la Computacin
Pontificia Universidad Católica de Chile
Av.Vicuña Mackenna 4860 (143), Santiago, Chile
Email: pmcortez@uc.cl, caundurr@uc.cl, dmery@ing.puc.cl, asoto@ing.puc.cl

***Abstract*—En visión por computador, la detección de objetos ha tenido un fuerte avance con la creación de nuevos descriptores de imagen. Un descriptor que ha aparecido recientemente es el descriptor de la matriz de covarianza, pero no se han hechos estudios sobre las diferentes metodologías para su construcción. Para resolver esta carencia hemos realizado un análisis sobre el aporte de distintas características de la imagen al descriptor y su aporte a la detección de diferentes objetos, en nuestro caso: caras y peatones. Es por ello que hemos definido un experimento con el cual determinar el desempeño de diferentes matrices de covarianza creadas a partir de distintas métricas de características. Con esto podemos determinar cuáles son las mejores sub características para los problemas de detección de objetos, rostros y peatones. También logramos destacar que no se puede utilizar cualquier tipo de combinación de sub características ya que puede que no exista una correlación entre ellas. Finalmente, al realizar un análisis con el mejor set de características, para el problema de detección de objetivos basado en un rostro obtuvimos un 99% de desempeño, mientras que para el problema de detección de objetivos basados en un peatón obtuvimos un 85% desempeño. Con esto esperamos tener una base más firme a la hora de elegir características para este descriptor y poder avanzar en otros tópicos como el reconocimiento o tracking de objetos.**

***Keywords*-Region Covariance, target detection.**

I. Introducción

Una de las habilidades más extraordinarias de la visión humana es el reconocimiento de objetos y rostros. Sin importar el ángulo, tamaño, luminosidad u oclusión del objeto, la visión humana logra en casi todos los casos, reconocer el objeto o persona. Esta habilidad es primordial en muchos aspectos de nuestras vidas, por ejemplo, sin la capacidad de reconocer rostros o expresiones faciales no podríamos tener una vida social satisfactoria. Teniendo en cuenta esta definición, el siguiente paso lógico será poder diseñar máquinas o siéstemas que puedan lograr esta habilidad automáticamente, para poder utilizarlos, por ejemplo, en aplicaciones de vigilancia o control de calidad. El área de visión por computador (o visión artificial) es un sub-campo del área de la inteligencia artificial, el objetivo global de éste es programar una máquina que logre entender o reconocer los patrones de una escena o las características de una imagen.

En el campo de visión por computador, lograr estas tareas es un desafío que todavía no se logra solucionar en su cabalidad. Gracias a los avances e investigaciones de los últimos años se han podido obtener múltiples aplicaciones de detección y reconocimiento en muchas áreas distintas. Estas incluyen video-juegos, asistencia para conductores, edición de video, control de calidad, control de tránsito, vigilancia, seguridad, *tracking*, etc. Por dar algunos ejemplos: en asistencia para conductores existen aplicaciones donde se le advierte al conductor si se está quedando dormido, basándose en reconocimiento de expresiones [1]; en control de calidad existen aplicaciones las cuales pueden definir si un producto está en perfecto estado o no, a partir de las características (tamaños, forma, etc.) de la imagen obtenida [2]; por último, en el área de seguridad y vigilancia existen aplicaciones las cuales, a partir del video de seguridad, detectan objetos extraños o comportamientos extraños (robos, violencia, etc.) [3].

Actualmente, para lograr estas tareas se utilizan distintas técnicas, a través de las cuales se obtiene información relevante de las imágenes o videos, conocidos como descriptores o características [4]. La selección de características es uno de los pasos más importantes en el problema de detección y reconocimiento. Un descriptor debe ser idealmente discriminativo, robusto y fácilmente computable. Existe una gran variedad de descriptores, algunos enfocados a ser calculados rápidamente, mientras que otros, en obtener la mayor información posible. Por otra parte existen algoritmos que detectan regiones relevantes e invariantes al tamaño, luminosidad y perspectiva, de esta forma se calculan las características solo a estas regiones relevantes y no a toda la imagen, esta tecnología se conoce en inglés como *viewpoint invariant segmentation* [5], [6].

En este trabajo hemos definido un experimento con el cual determinar el desempeño de diferentes matrices de covarianza creadas a partir de distintas sub características. Con esto podemos determinar cuáles son las mejores para los problemas de detección de objetos, rostros y peatones. Para esto, primero obtenemos un set de imágenes, donde se selecciona un objetivo específico que se desea detectar. A continuación obtenemos, en la imagen de búsqueda, la región de menor distancia a la región u objeto seleccionado

inicialmente. De esta forma, definimos un umbral de aceptación, que decide si el objetivo está o no, en la imagen de búsqueda y obtenemos el desempeño para cada set de sub características utilizadas en la formación del descriptor de covarianza. Finalmente, al realizar un análisis con el mejor set de características, para el problema de detección de objetivos basado en un rostro obtuvimos un 99% de desempeño, mientras que para el problema de detección de objetivos basados en un peatón obtuvimos un 85% desempeño.

Este artículo se organiza de la siguiente forma: en la sección 2 se describe el estado del arte actual del problema abordado; en la sección 3 se abordan las bases matemáticas, la hipótesis y la implementación del problema; a continuación, en la sección 4, se presenta la metodología y los resultados; finalmente, en la sección 5 se presentan las conclusiones.

II. Estado del Arte

En el área de detección de objetos existen distintos enfoques, uno de estos es el enfoque basado en características. En este enfoque se pueden distinguir dos procesos principales. La primera tarea es la extracción de características, las cuales deben otorgar la mayor información posible respecto al objeto, región o imagen. La segunda tarea es la detección del objeto o región a través de una buena clasificación de las características previamente obtenidas.

Los métodos de extracción de características pueden ser divididos en dos grupos, basados en su representación. El primer grupo de métodos es aquel que a partir de un algoritmo de detección de puntos relevantes se obtienen un conjunto de regiones locales representativas, por ejemplo: la detección de bordes y esquinas propuesto por Harris et al. en [7]; la detección por escala y relevancia en [8] o por regiones invariantes afines en [5] propuestas por Kadir et al. Métodos más recientes dejan de utilizar como descriptor la intensidad de la imagen y comienzan a utilizar los bordes y los gradientes de imágenes en un contexto espacial y a distintas escalas. Por ejemplo: el descriptor SIFT propuesto por Lowe en [9]; descriptores de contexto de la forma propuesto por Belongie et al. en [10]. Todos estos métodos basan su detección en establecer correspondencias entre los puntos de relevancia obtenidos de la imagen objetivo respecto de los extraídos de la imagen de origen. Muchos de estos algoritmos no son lo suficientemente robustos para la detección de peatones y caras, ya que no son invariantes a ciertas transformaciones como escalamiento y cambios de iluminación, dos grandes problemas a resolver. El más robusto de ellos ha demostrado ser SIFT, el cual es robusto a transformaciones planas, las cuales no son el caso de nuestro objetivo de detectar personas o caras.

El segundo grupo de métodos es aquel que encuentra un descriptor de objetos dentro de una ventana de detección. La imagen es densamente analizada buscando correspondencia entre las ventanas de origen y las ventanas de búsqueda. Estudios recientes utilizan como descriptor de objeto: plantillas de intensidad como los propuestos por Rowley et al. en [11] y Sung et al. en [12]; descriptores basados en Haar-Wavelets, los cuales son un set de funciones bases que codifican patrones visuales como las propuestas por Papageorgiou et al. en [13]. Estos métodos han sido bastante robustos para la detección de caras ya que la cantidad de deformaciones son pocas y bien conocidas. Por lo tanto podemos ver que ha sido completamente demostrado en éste contexto [14], [15], [16]. Pero en el problema de detectar elementos deformables se han visto pocas soluciones robustas. Es por esto que nos hemos visto en la necesidad de indagar más en ellos.

Recientemente Porikli et al. propusieron en [17] una elegante y simple solución para integrar múltiples características, las cuales son simples y rápidas de calcular; como gradiente, color, posición o intensidad, inclusive se pueden utilizar características de cámaras infrarrojas o térmicas. Este descriptor pertenece al segundo grupo de métodos descrito anteriormente, donde la región o ventana es representada por la matriz de covarianza de la matriz formada a partir de las características de la imagen. La región de covarianza se ha utilizado en distintas aplicaciones y se han propuesto diversas mejoras y complementos, por ejemplo: Tuzel et al. en [18] y Yao et al. en [19] proponen utilizar el descriptor de covarianza más un clasificador LogiBoost, para la detección de peatones; Hu et al. en [20] proponen utilizar el filtro partículas, para el *tracking* de objetos, utilizando como peso de las partículas, métricas de la matriz de covarianza; Meer et al. proponen en [21] un algoritmo para seguir objetos utilizando la región de covarianza y álgebra de Lie para crear un modelo de actualización. Todos estos innovadores aportes intentan mejorar el descriptor de covarianza, pero ninguna trata de relacionar la elección de sub características (como color, gradiente, etc.), con el problema a tratar. En nuestro trabajo aportaremos datos estadísticos sobre qué tipo de sub características es más útil dado el problema a tratar. De esta forma obtendremos la real implicancia de la selección de sub características en el descriptor.

III. Método Propuesto

A. Marco Teórico

El descriptor de covarianza propuesto por Porikli et al. en [17], se define formalmente como:

$$F(x, y, i) = \phi_i(I, x, y) \tag{1}$$

Donde I es una imagen (la cual puede estar en RGB, blanco y negro, infra-rojo, etc.), F es una matriz de $W \times H \times d$, donde W es el ancho de la imagen, H el alto de la imagen y d es el número de sub características utilizadas y ϕ_i es la función que relaciona la imagen con la i-ésima característica, es decir la función que obtiene la i-ésima características a partir de la imagen I. Es importante destacar que las características se obtienen a nivel del pixel (Figura 1).

El objetivo es representar el objeto a partir de la matriz de covarianza de la matriz F, construida a partir de estas características. La covarianza es la medición estadística de la variación o relación entre dos variables aleatorias, esta puede ser negativa, cero o positiva, dependiendo de la relación entre ellas. En nuestro caso las variables aleatorias representaran las sub características. En la matriz de covarianza las diagonales representan la varianza de cada característica, mientras que el resto representa la correlación entre las características.

Utilizar la matriz de covarianza como descriptor, tiene múltiples ventajas: 1) unifica información tanto espacial como estadística del objeto; 2) provee una elegante solución para fusionar distintas características y modalidades; 3) tiene una dimensionalidad muy baja; 4) es capaz de comparar regiones, sin estar restringido a un tamaño de ventana constante o fija, ya que no importa el tamaño de la región, el descriptor es la matriz de covarianza, que es de tamaño constante $d \times d$; 5) la matriz de covarianza puede ser fácilmente calculable, para cualquier región o sub-región.

A pesar de todos los beneficios que trae la representación del descriptor a partir de la matriz de covarianza, el cálculo para cualquier sub ventana o región dado una imagen, utilizando los métodos convencionales, la hace computacionalmente prohibitiva. Tuzel et al. en [22] proponen un método computacionalmente superior, para calcular la matriz de covarianza de cualquier sub ventana o región (rectangular) de una imagen a partir de la formulación de la imagen integral. El concepto de la imagen integral fue inicialmente introducida por Viola et al. en [23], para el cómputo rápido de características de Haar.

Sea P una matriz de $W \times H \times d$, el tensor de la imagen integral

$$P(x', y', i) = \sum_{x<x', y<y'} F(x, y, i) \quad i = 1 \dots d \tag{2}$$

Sea Q una matriz de $W \times H \times d \times d$, el tensor de segundo orden de la imagen integral

$$Q(x', y', i, j) = \sum_{x<x', y<y'} F(x, y, i) F(x, y, j) \tag{3}$$

$$i, j = 1 \dots d$$

Ahora, sea

$$P_{x,y} = \begin{bmatrix} P(x, y, 1) & \dots & P(x, y, d) \end{bmatrix}^T \tag{4}$$

$$Q_{x,y} = \begin{pmatrix} Q(x, y, 1, 1) & \dots & Q(x, y, 1, d) \\ \vdots & \ddots & \vdots \\ Q(x, y, d, 1) & \dots & Q(x, y, d, d) \end{pmatrix} \tag{5}$$

Hay que notar que la matriz $Q_{x,y}$ es simétrica y que para calcular P y Q se necesitan $d + (d^2 + d)/2$ pasos. La complejidad de calcular la imagen integral es de $O(d^2WH)$. En la Figura 2(a) vemos que la Matriz de Covarianza en un punto (x, y) representa la región desde el origen al punto dado. En la Figura 2(b) vemos gráficamente que la covarianza de cualquier región de la imagen se calcula como:

$$R_Q = Q_{x',y'} + Q_{x'',y''} - Q_{x'',y'} - Q_{x',y''} \tag{6}$$

$$R_P = P_{x',y'} + P_{x'',y''} - P_{x'',y'} - P_{x',y''} \tag{7}$$

$$C_{R(x',y';x'',y'')} = \frac{1}{n-1}[R_Q - \frac{1}{n} R_P R_P^T] \tag{8}$$

Donde $n = (x'' - x')(y'' - y')$. De esta forma, después de construir el tensor de primer orden P y el tensor de segundo orden Q, la covarianza de cualquier región se puede computar en $O(d^2)$.

Cabe destacar que, el descriptor de covarianza no es un elemento del espacio Euclidiano, por lo tanto no se pueden utilizar los clásicos algoritmos de inteligencia de máquinas, como por ejemplo: vecinos cercanos, distancia de Mahanalobis, etc. Por otra parte, las matrices de covarianza son simétricas positivas definidas, las cuales están incluidas dentro de la álgebra de Lie o la geometría de *Riemannian Manifolds* [18]. La álgebra Riemaniana manifold es un espacio topológico de manifold con métricas Riemaniana,

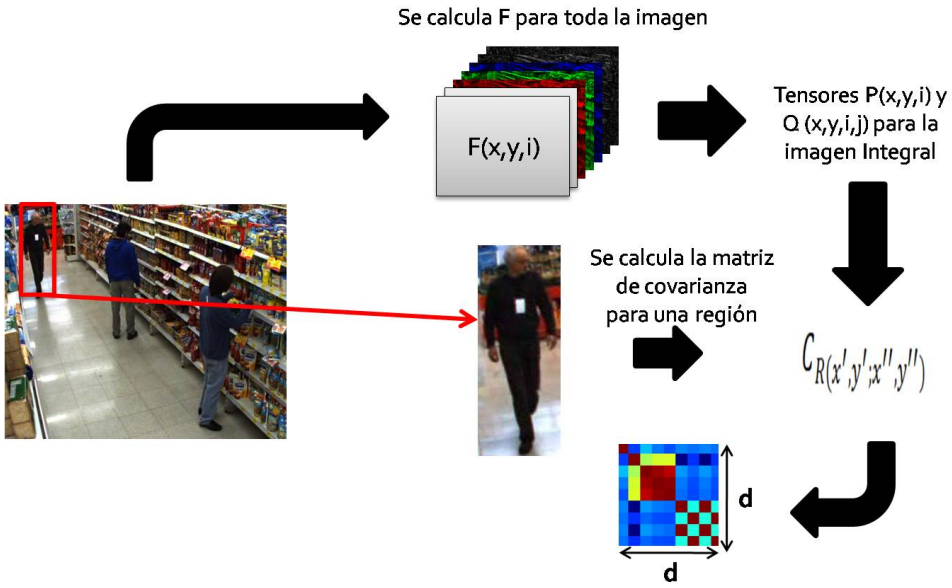


Figure 1. Ejemplo de cómo se construye el descriptor de covarianza de una región, a partir de una imagen pasando por la creación de la matriz de características.

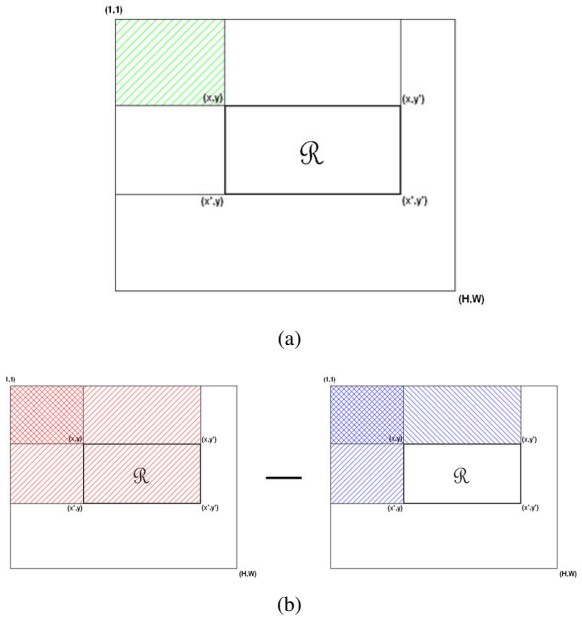


Figure 2. Representación gráfica del cálculo de la matriz de covarianza: (a) en un punto (x, y) dado; (b) para una región R dada a partir de los puntos (x, y, x', y').

las cuales permiten generalizar el espacio Euclidiano [24].

En nuestra investigación, para comparar dos regiones a partir de las matrices de covarianza, utilizaremos la métrica propuesta por Frstner y al. en [25]. La cual se define:

$$\rho(C_1, C_2) = \sqrt{\sum_{i=1}^{n} \ln^2 \lambda_i(C_1, C_2)} \tag{9}$$

Donde $\lambda_i(C_1, C_2)_{i=1..n}$ son los valores propios generalizados de C_1 y C_2 tal que,

$$\lambda_i C_1 x_i - C_2 x_i = 0 \quad i = 1 \dots d \tag{10}$$

Esta métrica satisface los axiomas de las matrices simétricas definidas positivas C_1 y C_2:

$$\rho(C_1, C_2) \geq 0 \tag{11}$$

$$\rho(C_1, C_2) = 0 \Rightarrow C_1 = C_2 \tag{12}$$

$$\rho(C_1, C_2) = \rho(C_2, C_1) \tag{13}$$

$$\rho(C_1, C_2) + \rho(C_1, C_3) \geq \rho(C_2, C_3) \tag{14}$$

Para crear las matrices utilizaremos diferentes espacios de colores los cuales proveen una poderosa información sobre el objeto a reconocer. Existen diferentes espacios de colores y se pueden inventar nuevos espacios con transformaciones de los ya existentes. El espacio más común es RGB (del inglés Red, Green, Blue) y de cual nacen varios otros, como el CMY, el cual se usa para televisión y no existe una conversión simple entre estos dos. Otros espacios son el HSL y HSV (del inglés Hue, Saturation, Lightness, Value)

los cuales se pueden obtener del espacio RGB.

Gevers et al. en [26] nos proponen los nuevos espacios $c_1c_2c_3$ y $l_1l_2l_3$. Por otra parte, también proponen el espacio $m_1m_2m_3$ el cual está definido como en relación a un pixel vecino. Nosotros proponemos utilizar el promedio de la vecindad. Sean $R^VG^VB^V$ los promedios de la vecindad del espacio RGB y $R^XG^XB^X$ los valores del pixel evaluado, todos los cuales quedan definidos como en la Tabla I y en la Tabla II se pueden apreciar sus invarianzas.

Table I
ECUACIONES PARA FORMAR LOS NUEVOS ESPACIOS DE COLORES.

Tabla de colores	
c_1	$\arctan(\frac{R}{max(G,B)})$
c_2	$\arctan(\frac{G}{max(R,B)})$
c_3	$\arctan(\frac{B}{max(R,G)})$
l_1	$\frac{(R-G)^2}{(R-G)^2+(R-B)^2+(G-B)^2}$
l_2	$\frac{(R-B)^2}{(R-G)^2+(R-B)^2+(G-B)^2}$
l_3	$\frac{(G-B)^2}{(R-G)^2+(R-B)^2+(G-B)^2}$
m_1	$\frac{R^XG^V}{R^VG^X}$
m_2	$\frac{R^XB^V}{R^VB^X}$
m_3	$\frac{G^XB^V}{G^VB^X}$

El objetivo de éste trabajo es utilizar estos siete espacios de colores, en diferentes problemas, para hacer un conjunto de pruebas experimentales.

B. Hipótesis

El problema definido en éste trabajo es encontrar un descriptor lo suficientemente eficiente, rápido de computar, y con altos grados de invariabilidad frente a diferentes condiciones de imágenes. El problema surge dado que los descriptores más invariantes son de mayor tamaño y por lo tanto tienen un mayor costo computacional.

De esta forma, deseamos demostrar que para distintos problemas de detección de un objeto se necesitan distintas sub características, para formar el descriptor de covarianza. De por sí el descriptor, al ser una matriz de covarianza, es invariante a ciertos cambios de iluminación y escala, pero depende profundamente de las sub características seleccionadas.

Finalmente nuestra hipótesis de trabajo ser demostrar que para distintos problemas, el rendimiento aumenta al utilizar distintas sub características para formar la matriz de covarianza.

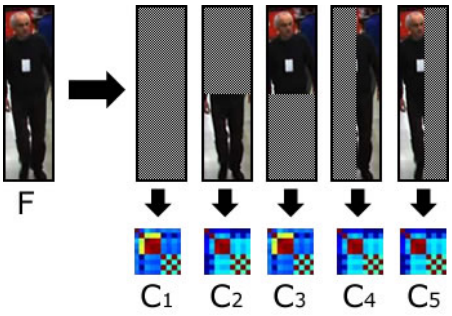


Figure 3. Distancia que disminuye la oclusión, asigna la distancia entre descriptores, como la menor distancia entre sub regiones.

C. Implementación

Para lograr los objetivos definidos previamente, en una primera parte, nos enfocaremos en implementar satisfactoriamente el descriptor de covarianza propuesto por Porikli et al. [17]. Esto incluye la implementación del nuevo método para el cálculo de la matriz de covarianza, para cualquier sub-región de una imagen propuesto por Porikli y Tuzel [22], la implementación de la distancia entre matrices de covarianza propuesto por Fröstner et al. [25] y la implementación de un algoritmo de búsqueda por ventanas dentro de la imagen (Figura 1). Toda esta implementación se realizo con el programa *MATLAB*.

Para implementar el descriptor de covarianza de una región, primero creamos la matriz F con (1), a continuación obtenemos los tensores de primer y segundo orden a partir de (2) y (3). Finalmente se obtiene el descriptor de covarianza de cualquier región a partir de (8). La idea es obtener la región con menor distancia entre descriptores de covarianza, en la imagen donde se está buscando el objeto.

En nuestra investigación utilizamos dos métricas distintas, para medir la similitud entre descriptores de covarianza. La primera es utilizar directamente la métrica basada en los valores propios generalizados, de dos matrices de covarianza, definida en (9) y la segunda (utilizada para deteccion de algun peatón) es utilizar una comparación de varios sub conjuntos de matriz de covarianza utilizando la distancia (9). La idea es disminuir la oclusión asignando la distancia entre descriptores, como la menor distancia entre los descriptores de cada sub-región (Figura 3).

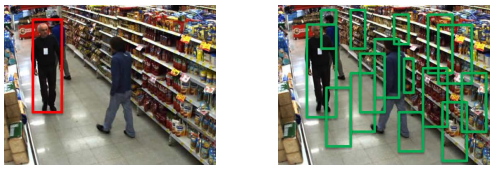
$$\rho(O,T) = \min_j \rho(C_j^O, C_j^T) \tag{15}$$

Donde C_j^O es la matriz de covarianza de la sub región, de la región origen y C_j^T es la matriz de covarianza de la sub región, de la región de búsqueda

A continuación, para hacer el detección del objeto en la imagen de búsqueda, utilizamos un algoritmo de fuerza

Table II
RESUMEN DE LOS DISTINTOS ESPACIOS DE COLORES Y SUS INVARIANCIAS A VARIAS CONDICIONES (+ DENOTA INVARIANTE - DENOTA SENSIBLE A LA CONDICIÓN) [26].

	viewing direction	surface orientation	highlights	illumination direction	illumination intensity	illumination color	inter reflection
I	-	-	-	-	-	-	-
RGB	-	-	-	-	-	-	-
rgb	+	+	-	+	+	-	-
S	+	+	-	+	+	-	-
$c_1c_2c_3$	+	+	-	+	+	-	-
H	+	+	+	+	+	-	-
$l_1l_2l_3$	+	+	+	+	+	-	-
$m_1m_2m_3$	+	+	-	+	+	+	+



(a) Imagen Origen (b) Imagen Búsqueda

Figure 4. Representación de la imagen de origen y búsqueda: (a) Imagen de origen y la región de origen en rojo; (b) Imagen de búsqueda y todas las regiones de búsqueda en verde.

bruta (Figura 4), ya que teniendo calculado los tensores de primer y segundo orden de una imagen, podemos calcular el descriptor de covarianza para cualquier sub regiones, en $O(d^2)$. De esta forma, primero se compara, en la imagen de búsqueda, con 300 regiones o ventanas aleatorias del tamaño de la región de origen. De estas 300 regiones se seleccionan las 4 con menor distancia a la región de origen. A partir de cada una de las 4 regiones seleccionadas, se busca aleatoriamente en 30 regiones con distintos tamaños. Cada una de stas nuevas regiones tiene su centro dentro de la región seleccionada inicialmente. Luego se selecciona la región con menor distancia, a la región de origen, de las 120 regiones de búsqueda.

Finalmente, utilizamos ocho distintas matrices F (basados en los colores propuestos en la sección anterior), para formar los descriptores de covarianza. Las matrices F definidas se pueden observar en la Tabla III. Donde R, G y B son rojo (Red), verde (Green) y azul (Blue); $|I_x|$ es la primera derivada de la intensidad en la dirección x, $|I_y|$ es la primera derivada de la intensidad en la dirección y; $|I_{xx}|$ es la segunda derivada de la intensidad en la dirección x; $|I_{yy}|$ es la segunda derivada de la intensidad en la dirección y; $tan^{-1}(\frac{I_x}{I_y})$ corresponde a las orientaciones de los bordes.

IV. EXPERIMENTOS Y RESULTADOS

A. Metodología

Antes de comenzar a medir el desempeño del descriptor, se debe definir inicialmente la metodología de prueba. Para esto primero se debe seleccionar la región que se desea detectar en la imagen origen, llamemos esta región, región origen. A continuación se busca la región más parecida o de menor distancia a la región origen en otra imagen (Figura 4). La región encontrada se llama la región objetivo o de búsqueda.

Para poder saber si una imagen tiene, o no, la región buscada, definimos un factor k, el cual define un límite o umbral de aceptación de la distancia medida entre los descriptores de covarianza. Por lo tanto, si la distancia entre dos descriptores de covarianza es mayor a k, la imagen no tiene el objeto buscado, mientras que sí es menor o igual, si lo tiene. Por otra parte, si al clasificar se obtiene una distancia menor que el factor k, pero la región objetivo está mal ubicada, consideraremos este caso como falso positivo.

Para obtener todos los resultados se utilizarán dos videos de 640×480, filmados a 30 cuadros por segundo, en un supermercado local (Santiago, Chile) con cámaras Point Grey (Figura 5). El primer video (video góndolas) se utilizar para la detección de un objeto o peatón dado, mientras que el segundo video (video cajas) se utilizar para el detección de una cara u objeto dado. Esta diferencia se hace ya que el primer video no tiene la resolución adecuada para la detección de rostros. Finalmente, a partir de estos videos se obtendrán dos sets de 200 imágenes, para hacer la detección del objeto u rostro, donde en 100 se encuentra el objeto seleccionado inicialmente y en las otras 100 no se encuentra.

B. Resultados

Los resultados siguientes describen el desempeño del descriptor de covarianza, a partir de las distintas matrices F anteriormente definidas, para cada uno de los sets de imagenes.

Cabe destacar que las características F_2 y F_3 dieron desempeños exactamente iguales, la características F_8 no otorgaba suficiente información y por lo tanto no se pudo calcular correctamente la matriz de covarianza y las características de la matriz F_9 es un conjunto de todos los

Table III
CARACTERÍSTICAS PARA FORMAR LAS MATRICES F.

Matrices F - Sub Características	
F_1	$\begin{bmatrix} x & y & R & G & B & \lvert I_x\rvert & \lvert I_y\rvert & \lvert I_{xx}\rvert & \lvert I_{yy}\rvert \end{bmatrix}$
F_2	$\begin{bmatrix} x & y & H & S & L & \lvert L_x\rvert & \lvert L_y\rvert & \lvert L_{xx}\rvert & \lvert L_{yy}\rvert \end{bmatrix}$
F_3	$\begin{bmatrix} x & y & H & S & V & \lvert V_x\rvert & \lvert V_y\rvert & \lvert V_{xx}\rvert & \lvert V_{yy}\rvert \end{bmatrix}$
F_4	$\begin{bmatrix} x & y & R & G & B & \lvert I_x\rvert & \lvert I_y\rvert & \sqrt{\lvert I_x\rvert^2+\lvert I_y\rvert^2} & \lvert I_{xx}\rvert & \lvert I_{yy}\rvert & \tan^{-1}(\frac{I_x}{I_y}) \end{bmatrix}$
F_5	$\begin{bmatrix} x & y & \lvert I_x\rvert & \lvert I_y\rvert & \sqrt{\lvert I_x\rvert^2+\lvert I_y\rvert^2} & \lvert I_{xx}\rvert & \lvert I_{yy}\rvert & \tan^{-1}(\frac{I_x}{I_y}) \end{bmatrix}$
F_6	$\begin{bmatrix} x & y & c_1 & c_2 & c_3 & \lvert I_x\rvert & \lvert I_y\rvert & \lvert I_{xx}\rvert & \lvert I_{yy}\rvert \end{bmatrix}$
F_7	$\begin{bmatrix} x & y & l_1 & l_2 & l_3 & \lvert I_x\rvert & \lvert I_y\rvert & \lvert I_{xx}\rvert & \lvert I_{yy}\rvert \end{bmatrix}$
F_8	$\begin{bmatrix} x & y & m_1 & m_2 & m_3 & \lvert I_x\rvert & \lvert I_y\rvert & \lvert I_{xx}\rvert & \lvert I_{yy}\rvert \end{bmatrix}$
F_9	$\begin{bmatrix} x & y & R & G & B & H & S & L & c_1 & c_2 & c_3 & l_1 & l_2 & l_3 & \lvert I_x\rvert & \lvert I_y\rvert & \sqrt{\lvert I_x\rvert^2+\lvert I_y\rvert^2} & \lvert I_{xx}\rvert & \lvert I_{yy}\rvert & \tan^{-1}(\frac{I_x}{I_y}) \end{bmatrix}$

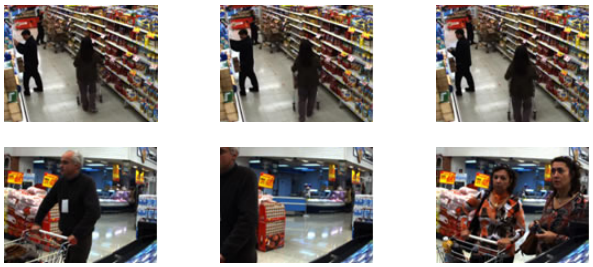


Figure 5. Video 1 - góndolas: detección de un peatón u objeto (imagen superior). Video 2 - Caja: detección de un rostro u objetos (imagen inferior).

espacios de colores utilizados (excepto el $m_1m_2m_3$), más las derivadas de la intensidad y la orientación de los bordes; de esta forma podremos observar si en conjunto los espacios de colores otorgan mayor información o correlación, que por separado. Por otra parte, para poder comparar los resultados se normalizaron todas las distancias calculadas entre los descriptores de covarianza. Los resultados se encuentran resumidos en la Tabla IV y en las Figuras 6(a) y 6(b) se encuentran las Curvas ROC para distintos factores k.

Table IV
RENDIMIENTO PARA TODAS LAS MATRICES F, UTILIZANDO EL MEJOR FACTOR k.

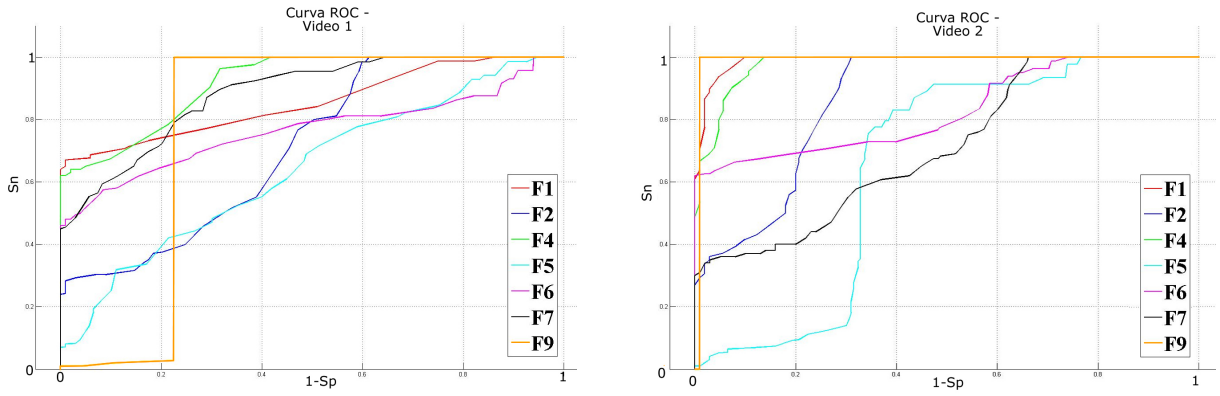
Rendimiento		
Caracterstica	Video1	Video2
F_1	94%	80%
F_2	78%	64%
F_4	92%	83%
F_5	68%	63%
F_6	81%	76%
F_7	66%	78%
F_9	**99%**	**85%**

C. *Análisis*

A partir de los resultados obtenidos podemos afirmar que, para el problema de detección de una cara específica, las mejores sub características son las definidas por F_9, con un 99% de rendimiento y F_1, con un 94% de rendimiento, mientras que, para el problema de detección de un peatón u objeto especifico, las mejores sub características son las definidas por F_9, con un 85% de rendimiento y F_4, con un rendimiento de 83%. También hay que destacar que, los resultados muestran que las sub características relacionadas con los colores, son muy importante, independiente del problema, sobre todo en el caso de RGB, ya que las sub características de la matriz F_5, que no incluyen ningún espacio de color, obtuvieron, en los dos casos, los peores desempeños.

De esta forma, podemos observar que el conjunto de sub características de la matriz F_9 (un conjunto de las matriz F utilizadas), obtienen mayor rendimiento y por lo tanto otorgan mayor información o correlación, que cada una de las matrices F por separada. Este resultado es esperable, ya que al utilizar este conjunto de sub características, utilizamos todas las correlaciones posibles, entre pares de sub características. Lamentablemente, utilizar una matriz F de tantas dimensiones, hace que el computo de los tensiones P y Q sea prohibitivo para cierto tamaño de imagen.

Por otra parte las sub características obtenidas por F_7, muestran un desempeño alto, pero no se obtuvo gran discernimiento entre imágenes que tenían, o no, el objeto, produciendo gran cantidad de falsos positivos. También hay que destacar que las sub características de F_1 y F_4 son muy similares, pero la sub característica $\tan^{-1}(\frac{I_x}{I_y})$ otorga gran cantidad información relevante. Lamentablemente, F_4 al tener mayor cantidad de sub características, tiene mayor dimensionalidad, provocando un mayor tiempo de cálculo



(a) Set imagenes: Peatones

(b) Set imagenes: Rostros

Figure 6. Curvas ROC para las características F_1 a F_7

para los tensores. Se pueden observar más ejemplos en las Figuras 7, 8, 9 y 10, donde en cada set imagenes se encuentra el objetivo seleccionado inicialmente. Podemos observar que al seleccionar inicialmente rostros u objetos, la detección es bastante precisa, mientras que cuando se selecciona un peatón, el rendimiento disminuye. Finalmente cabe destacar que las sub características de F_4 tuvieron un gran desempeño, pero con menor rendimiento, ya que tienen dificultades, para detectar si el objeto está o no en la imagen, dando as falsos positivos.

Cabe destacar, que todas las pruebas se realizaron en un computador con un procesador Intel Core 2 Duo y 2 Gb en ram. De esta forma, los tiempos de ejecución fueron del orden de 6 a 7 segundos por imagen procesada, estos tiempos son directamente proporcionales al tamaño de la imagen.

V. Conclusiones y Trabajos futuros

A partir de los estudios presentados podemos afirmar, que el descriptor de covarianza es robusto a cambios de iluminación y formas, pero tiene cierta debilidad en los cambios de escala, lo cual posiblemente puede ser resuelto por una normalización de la matriz de covarianza. Cabe destacar la importancia de utilizar los colores como sub características, especialmente el espacio RGB ya que tiene una gran correlación con los gradientes de intensidad de la imagen.

Las mejores sub características, para los problemas de deteccion de un objeto, rostro o peatón especifico son las de F_9, F_1 y F_4. Donde es importante recordar que, la matriz de sub características F_9 es un conjunto de todas las sub características utilizadas en las otras matrices F y tiene un tiempo mayor de calculo. Ésto demuestra que es importante la selección de características, para disminuir el tiempo de cálculo, y como se relacionan entre ellas, para mejorar su rendimiento. Es importante destacar que no se puede utilizar cualquier tipo de combinación de sub características, ya que al momento de realizar la matriz de covarianza, puede que no exista suficiente correlación entre las sub características, como se observó con F_8.

Hay que destacar que el descriptor tiene gran futuro ya que logra unificar tanto información espacial como estadísticas. Es por esto, que se continuará trabajando en disminuir tiempo de ejecución, en construir una metodología completa, para obtener las características más relevantes para un problema dado y en una aplicación de disminución de dimensiones para tensores, utilizando MPCA, para la matriz F. Todo esto es un trabajo preparatorio para implementar un método eficiente e innovador de tracking.

Agradecimientos

Agradecemos a Fernando Betteley de Cencosud por facilitar las instalaciones de Supermercados Santa Isabel para la adquisición de videos. Esta investigación ha sido financiada en parte por LACCIR (Latin American and Caribbean Collaborative ICT Research).

References

[1] Q. Ji, Z. Zhu, and P. Lan, "Real-time nonintrusive monitoring and prediction of driver fatigue," *IEEE transactions on vehicular technology*, vol. 53, no. 4, 2004.

[2] F. P. Edreschi, D. Mery, F. Mendoza, and J. M. Aguilera, "Classification of potato chips using pattern recognition," *Journal of Food Science*, vol. 69, no. 6, pp. 264–270, 2004.

[3] N. T. Nguyen, H. H. Bui, S. Venkatsh, and G. West, "Recognizing and monitoring high-level behaviors in complex spatial environments," in *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings*, vol. 2, 2003.

[4] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.

[5] T. Kadir, A. Zisserman, and M. Brady, "An affine invariant salient region detector," *Lecture Notes in Computer Science*, pp. 228–241, 2004.

[6] J. Sivic, F. Schaffalitzky, and A. Zisserman, "Efficient object retrieval from videos," in *12th European Signal Processing Conference (EUSIPCO'04)*, 2004.

[7] C. Harris and M. Stephens, "A combined corner and edge detector," in *Alvey vision conference*, vol. 15, 1988, p. 50.

[8] T. Kadir and M. Brady, "Saliency, scale and image description," *International Journal of Computer Vision*, vol. 45, no. 2, pp. 83–105, 2001.

[9] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[10] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 509–522, 2002.

[11] H. A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," in *1996 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1996. Proceedings CVPR'96*, 1996, pp. 203–208.

[12] K. K. Sung and T. Poggio, "Example-based learning for view-based human face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 39–51, 1998.

[13] C. Papageorgiou and T. Poggio, "A trainable system for object detection," *International Journal of Computer Vision*, vol. 38, no. 1, pp. 15–33, 2000.

[14] S. Baker, I. Matthews, J. Xiao, R. Gross, T. Kanade, and T. Ishikawa, "Real-time non-rigid driver head tracking for driver mental state estimation," in *11th World Congress on Intelligent Transportation Systems*. Citeseer, 2004.

[15] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, 2001.

[16] D. DeCarlo and D. Metaxas, "Deformable model-based shape and motion analysis from images usingmotion residual error," in *Computer Vision, 1998. Sixth International Conference on*, 1998, pp. 113–119.

[17] O. Tuzel, F. Porikli, and P. Meer, "Region covariance: A fast descriptor for detection and classification," *Lecture Notes in Computer Science*, vol. 3952, p. 589, 2006.

[18] ——, "Pedestrian detection via classification on riemannian manifolds," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 10, pp. 1713–1727, 2008.

[19] J. Yao and J. M. Odobez, "Fast human detection from videos using covariance features," in *ECCV 2008 Visual Surveillance Workshop*, 2008.

[20] H. Hu, J. Qin, Y. Lin, and Y. Xu, "Region covariance based probabilistic tracking," in *Intelligent Control and Automation, 2008. WCICA 2008. 7th World Congress on*, 2008, pp. 575–580.

[21] F. Porikli, O. Tuzel, and P. Meer, "Covariance tracking using model update based on lie algebra," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, 2006.

[22] F. Porikli and O. Tuzel, "Fast construction of covariance matrices for arbitrary size image windows," in *Proc. Intl. Conf. on Image Processing*, 2006, pp. 1581–1584.

[23] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple," in *Proceedings of CVPR2001*, vol. 1, 2001.

[24] W. Rossmann, *Lie Groups: An introduction Through Linear Goups.* Oxford Press, 2002.

[25] W. Forstner and B. Moonen, "A metric for covariance matrices," *Qua vadis geodesia*, pp. 113–128, 1999.

[26] T. Gevers and A. W. M. Smeulders, "Color-based object recognition," *Pattern recognition*, vol. 32, no. 3, pp. 453–464, 1999.

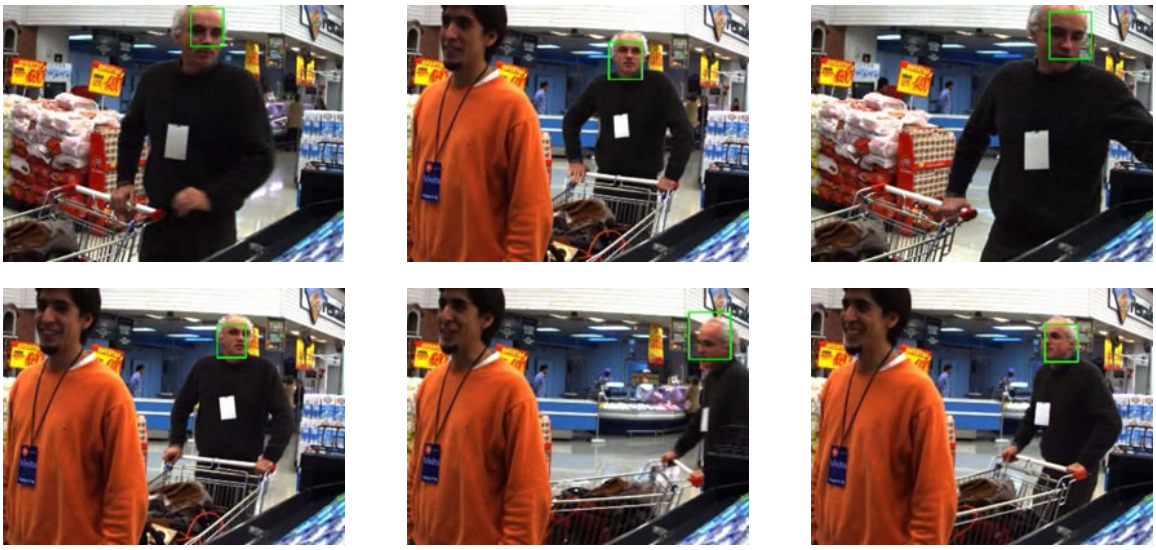


Figure 7. Detección de un rostro dado (rectángulo verde), con sub características de F_4

Figure 8. Detección de un objeto dado (rectángulo verde), con sub características de F_4

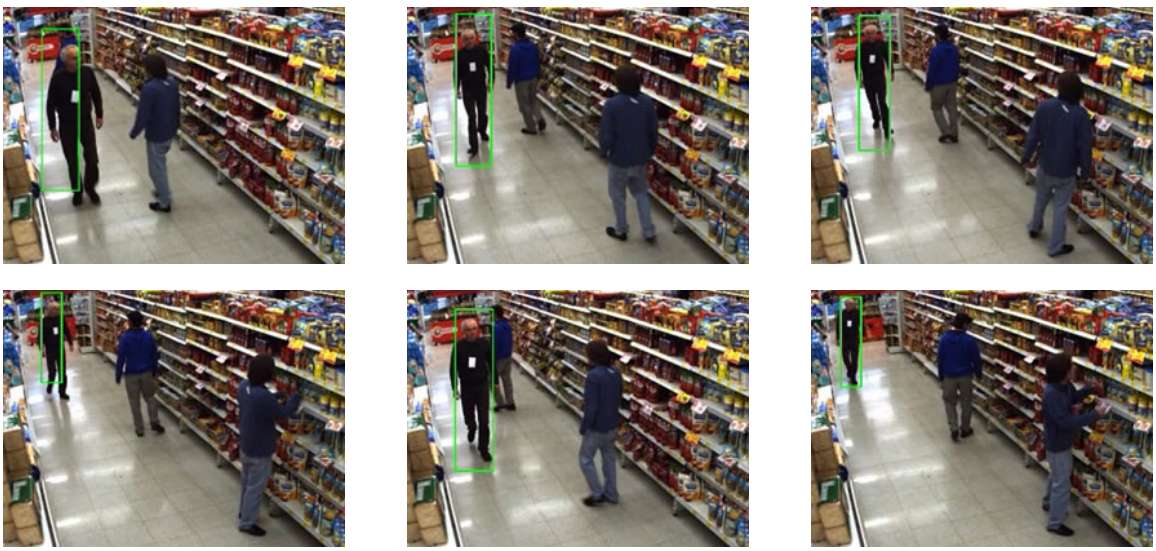


Figure 9. Detección de un peatón dado (rectángulo verde), con sub características de F_4

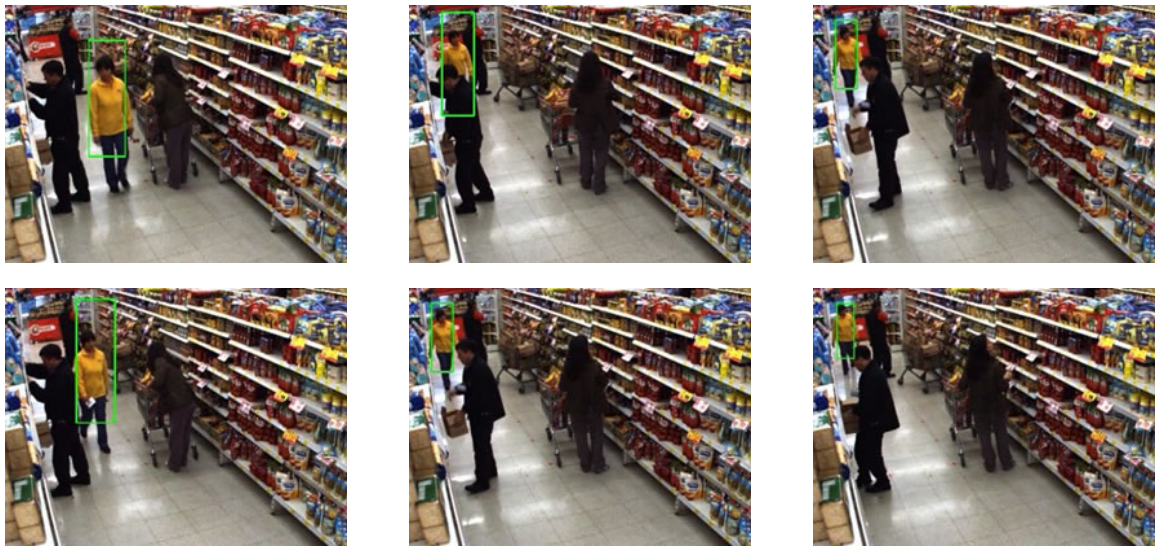


Figure 10. Detección de un peatón dado (rectángulo verde), con sub características de F_4