

COMPARACIÓN Y DETECCIÓN DE PATRONES DE MOVIMIENTO EN VIDEO

ANDRÉS ESTEBAN PUERTO LARA
JAIME GACHARNÁ MADRIGAL

BOGOTÁ D.C.
PONTIFICIA UNIVERSIDAD JAVERIANA
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE ELECTRÓNICA
NOVIEMBRE DE 2010

COMPARACIÓN Y DETECCIÓN DE PATRONES DE MOVIMIENTO EN VIDEO

TRABAJO DE GRADO N° 0959

PROYECTO

ANDRÉS ESTEBAN PUERTO LARA

JAIME GACHARNÁ MADRIGAL

DIRECTOR ING. JULIAN QUIROGA SEPULVEDA, M.Sc.

BOGOTÁ D.C.

PONTIFICIA UNIVERSIDAD JAVERIANA

FACULTAD DE INGENIERÍA

DEPARTAMENTO DE ELECTRÓNICA

NOVIEMBRE DE 2010

PONTIFICIA UNIVERSIDAD JAVERIANA
FACULTAD DE INGENIERÍA

DEPARTAMENTO DE ELECTRÓNICA

RECTOR MAGNÍFICO:	JOAQUÍN SÁNCHEZ GARCÍA S. J.
DECANO ACADÉMICO:	Ing. FRANCISCO J. REBOLLEDO M.
DECANO DEL MEDIO UNIVERSITARIO:	SERGIO BERNAL RESTREPO S. J.
DIRECTOR DE CARRERA:	Ing. JUAN MANUEL CRUZ
DIRECTOR DEL TRABAJO DE GRADO:	Ing. JULIÁN QUIROGA SEPÚLVEDA

ARTÍCULO 23 DE LA RESOLUCIÓN N° 13 DE JUNIO DE 1946

“La universidad no se hace responsable de los conceptos emitidos por sus alumnos en sus trabajos de grado. Solo velará porque no se publique nada contrario al dogma y la moral católica y porque los trabajos no contengan ataques o polémicas puramente personales. Antes bien que se vea en ellos el anhelo de buscar la verdad y la justicia”

Tabla de Contenido

1.	INTRODUCCIÓN	7
2.	MARCO TEÓRICO.....	9
2.1	Video	9
2.2	Gradiente para una secuencia de video	9
2.3	Rango de una matriz.....	11
2.4	Matriz de rango deficiente.....	11
2.5	Valores propios de una matriz.....	11
2.6	Vectores propios de una matriz.....	11
2.7	Norma de Frobenius	11
2.8	Entrelazar valores propios en matrices simétricas	12
2.9	Matriz de Gram	12
2.10	Tensor de estructura.....	12
3.	DESCRIPCIÓN GENERAL Y ESPECIFICACIONES	13
3.1	Equipo empleado.....	13
3.2	Condiciones de los videos	13
3.3	Clasificación de los videos.....	14
3.4	Diagrama en bloques del sistema.....	14
3.5	Diagrama de flujo.....	16
4.	DESARROLLO TEÓRICO	20
4.1	Segmentación y puntaje de correlación.....	21
4.1.1	Obtención de sub-plantillas de V	21
4.1.2	Diezmado y filtrado de la plantilla T y la sub-plantilla S (pre-procesamiento)	22
4.1.3	Puntaje de correlación de la sub-plantilla S	23

4.1.4	Detección mayor puntaje de correlación (modificación a [5])	37
4.2	Creación del Video Resultado R (modificación a [5])	39
5.	ANÁLISIS DE RESULTADOS	41
5.1	Pruebas de Calibración.....	41
5.1.1	Comparación de la plantilla T con ella misma.....	44
5.1.2	Desplazamiento espacial por parches alrededor de la plantilla	44
5.1.3	Desplazamiento espacial por pixeles alrededor de la plantilla	46
5.1.4	Desplazamiento temporal alrededor de la plantilla.....	47
5.1.5	Búsqueda de un movimiento IDENTICO al de la plantilla	50
5.1.6	Búsqueda de un movimiento SIMILAR al de la plantilla.....	51
5.1.7	Sensibilidad a la escala	53
5.1.8	Búsqueda en videos sin un movimiento similar al de la plantilla.....	55
5.2	Pruebas Generales	57
5.2.1	Prueba No. 1	59
5.2.2	Prueba No. 2	60
5.2.3	Prueba No. 3	60
5.2.4	Prueba No. 4	61
5.2.5	Prueba No. 5	62
5.2.6	Análisis de pruebas generales	62
6.	CONCLUSIONES	65
7.	BIBLIOGRAFÍA.....	68
8.	ANEXOS.....	69
8.1	Anexo 1: Código en C++	69
8.2	Anexo 2: Carpeta de resultados.....	69

1. INTRODUCCIÓN

La detección de patrones de movimiento es una herramienta muy útil en diversos campos del conocimiento y de la industria, ya sea para estudios de comportamiento, seguridad o incluso el diagnóstico de diferentes enfermedades. Es el caso, por ejemplo, de la seguridad y la movilidad en las ciudades que presentan un alto tráfico vehicular, donde la detección no supervisada de movimientos no permitidos por parte de los vehículos es de gran utilidad para proporcionar una señal de advertencia al personal de seguridad. De esta manera, la detección de infracciones vehiculares, como conductores realizando algún cruce o giro en U prohibido, serían mucho más factibles en ciudades donde el parque automotor es tan elevado.

Del mismo modo, la comparación de patrones de movimiento es supremamente útil en aplicaciones deportivas, donde el estudio de los movimientos precisos y adecuados, se hace primordial para la optimización de la práctica de cada deporte [1]. Igualmente, la evaluación de algunos movimientos no permitidos en ciertas disciplinas deportivas se facilita y se precisa con un análisis sistemático por computadora, reduciendo la responsabilidad de la subjetividad en los jueces.

Existen una gran variedad de métodos y algoritmos para realizar un análisis del movimiento descrito en una secuencia de video. Por un lado, la estimación del fondo en una secuencia espacio-temporal (requerida usualmente en métodos de acción-reconocimiento [2]), implica una alta sensibilidad al ruido y a cambios en la iluminación, además de requerir una configuración ad-hoc. Por otro lado, algoritmos que requieren el aprendizaje previo de actividades como aquellos que utilizan redes neuronales están restringidos, en su aplicación, a un pequeño conjunto de actividades predefinidas (redes supervisadas) [3], [4]. Finalmente, algoritmos que realizan seguimiento y estimación explícita del movimiento, están expuestos a los obstáculos fundamentales del flujo óptico como los problemas de apertura.

Por tal motivo y en busca de nuevas alternativas, en este trabajo de grado se implementó una versión en C++ del algoritmo propuesto por Eli Shechtman y Michal Irani en [5], en el que por medio de la aplicación de conceptos matemáticos tridimensionales, se logra detectar y comparar movimientos determinados, sin necesidad de encontrar o estimar el objeto de manera específica.

Debido a la manera en que se desarrolla el algoritmo implementado, más que objetos de interés lo que nos interesa son movimientos específicos, pues el algoritmo pretende encontrar movimientos y no objetos; lo que proporciona un enfoque diferente al tratar de detectar y comparar patrones de movimiento en video.

Específicamente, en este trabajo de grado se analizan secuencias de video que contienen automóviles realizando giros en L y personas practicando deportes con movimientos complejos; siendo estos los patrones de movimiento específicos que nos interesa evaluar con este trabajo, sin descartar otro tipo de aplicaciones, por lo cual se realizó un implementación flexible del algoritmo.

2. MARCO TEÓRICO

Para abordar la detección de movimientos específicos desde un punto de vista enteramente matemático, como lo sugiere [5], es necesaria una pequeña introducción conceptual como se muestra a continuación.

2.1 Video

Un video se puede considerar como una secuencia de imágenes o fotogramas que varían en el tiempo. Se define como una función $f_c(x, y, t)$ que depende de dos variables espaciales (x, y) y una temporal t , usualmente observada en una ventana rectangular W sobre algún intervalo de tiempo T .

Las dimensiones de W se denotan como $m \times n$ donde m es el ancho de la ventana y n el alto, correspondiente a cada fotograma perteneciente al video [6].

2.2 Gradiente para una secuencia de video

El gradiente es un vector que indica en cada punto la dirección de máximo incremento de una función de múltiples variables. Sea P un pixel perteneciente a un video V y a su vez es una función de las variables (x, y, t) ; se define el gradiente de esta función como:

$$\nabla P_v = (P_{vx}, P_{vy}, P_{vt})$$

Dónde $(x, y, y t)$ representan la posición de un pixel en cualquier momento del video. Se definen las derivadas respecto a cada variable de la siguiente manera:

$$\begin{aligned} P_{vx} &= \frac{\partial P_v(x, y, t)}{\partial x} \\ P_{vy} &= \frac{\partial P_v(x, y, t)}{\partial y} \\ P_{vt} &= \frac{\partial P_v(x, y, t)}{\partial t} \end{aligned} \quad (1)$$

Para detectar los cambios de niveles de gris en un fotograma del video, se utiliza el operador derivada ya que éste es sensible a este tipo de cambios. Se debe tener en cuenta que una imagen

es un arreglo de dos dimensiones y por tanto se deben considerar las derivadas tanto en x como en y . Para esto se utiliza la función gradiente:

Sea I un fotograma perteneciente a V se definen las derivadas espaciales de I respecto a x y respecto a y de la siguiente manera:

$$\nabla I_v(x, y) = \left(\frac{\partial I_v}{\partial x}, \frac{\partial I_v}{\partial y} \right) \quad (2)$$

Como una imagen no es una función continua, se utilizan las aproximaciones por diferencias:

$$\frac{\partial I_v}{\partial x} = I_v(x+1, y) - I_v(x-1, y) \quad (3)$$

$$\frac{\partial I_v}{\partial y} = I_v(x, y+1) - I_v(x, y-1) \quad (4)$$

Estas expresiones son las derivadas alrededor de un pixel ubicado en la posición (x, y) .

Al extender el problema a un conjunto de imágenes en el tiempo (video), entonces también se deben considerar las derivadas temporales. Se define el gradiente espacio temporal de un pixel perteneciente a un video V que depende de (x, y, t) de la siguiente manera:

Haciendo la aproximación por diferencias, se tiene:

$$\frac{\partial P_v}{\partial x} = P_v(x+1, y, t) - P_v(x-1, y, t) \quad (5)$$

$$\frac{\partial P_v}{\partial y} = P_v(x, y+1, t) - P_v(x, y-1, t) \quad (6)$$

$$\frac{\partial P_v}{\partial t} = P_v(x, y, t+1) - P_v(x, y, t-1) \quad (7)$$

Estas expresiones son las derivadas alrededor de un pixel ubicado en la posición (x, y, t) .

2.3 Rango de una matriz

El rango de una matriz A , denotado por $\text{rango}(A)$, es la dimensión del espacio de columnas de A , es decir el número de columnas pivote de A cuando ésta está en la forma escalonada.

De una manera más formal, el rango de una matriz A de tamaño $m \times n$ se define como:

$$\text{rango}(A) = \dim(\{y \in \mathfrak{R}^m : y = Ax \text{ Para algún } x \in \mathfrak{R}^n\})$$

2.4 Matriz de rango deficiente

Se dice que una matriz $A \in \mathfrak{R}^{m \times n}$ es de rango deficiente si $\text{rango}(A) < \min\{m, n\}$

2.5 Valores propios de una matriz

Los valores propios de una matriz $A \in C^{n \times n}$ son las n raíces de su polinomio característico

$p(z) = \det(zI - A)$. El conjunto de estas raíces es llamado espectro y es denotado por $\lambda(A) = \{\lambda_1, \dots, \lambda_n\}$, de esto se sigue que $\det(A) = \lambda_1 \lambda_2 \dots \lambda_n$.

2.6 Vectores propios de una matriz

Un vector propio de una matriz A $n \times n$ es un vector x diferente de cero tal que $Ax = \lambda x$ para algún escalar λ .

Básicamente un vector propio define un sub-espacio unidimensional que es invariante con respecto a la pre-multiplicación por A . De manera más general, un sub-espacio $S \subseteq C^n$ con la siguiente propiedad:

$$x \in S \Rightarrow Ax \in S$$

se dice que es invariante (para A)

2.7 Norma de Frobenius

Sea A una matriz de $m \times n$, la norma de Frobenius de A está definida como:

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2}$$

$$a_{ij} \in A$$

$$i, j = 1, 2, \dots, n$$

2.8 Entrelazar valores propios en matrices simétricas

Si $A \in \mathfrak{R}^{n \times n}$ es simétrica y $A_r = A(1:r, 1:r)$ entonces:

$$\lambda_{r+1}(A_{r+1}) \leq \lambda_r(A_r) \leq \lambda_r(A_{r+1}) \leq \dots \leq \lambda_2(A_{r+1}) \leq \lambda_1(A_r) \leq \lambda_1(A_{r+1})$$

Para $r = 1: n-1$

2.9 Matriz de Gram

La matriz de Gram de un conjunto de vectores v_1, \dots, v_n , es la matriz que define el producto escalar, cuyas entradas vienen dadas por $G_{ij} = (v_i | v_j)$

Propiedades:

1. Simetría: $g_{ij} = g_{ji}$
2. Es una matriz semidefinida positiva
3. El primer elemento es positivo: $g_{11} > 0$
4. Su determinante es positivo.

2.10 Tensor de estructura

Conocido también como la matriz de segundo momento, es una matriz que contiene las derivadas del gradiente de una función de n variables. En esta matriz se resumen las direcciones predominantes en la vecindad de un punto. A saber, el tensor de estructura en un espacio 3D para un conjunto de 3 fotogramas P perteneciente a un video V es:

$$M[p] = \begin{bmatrix} (P_x[p])^2 & P_x[p]P_y[p] & P_x[p]P_t[p] \\ P_x[p]P_y[p] & (P_y[p])^2 & P_y[p]P_t[p] \\ P_x[p]P_t[p] & P_y[p]P_t[p] & (P_t[p])^2 \end{bmatrix}$$

3. DESCRIPCIÓN GENERAL Y ESPECIFICACIONES

Este trabajo de grado tiene como finalidad implementar una versión del algoritmo propuesto por Eli Shechtman y Michal Irani en [5], que se utiliza para definir si dos patrones de intensidad espacio-temporal diferentes que hacen parte de dos segmentos de video distintos podrían haber resultado de un campo similar de movimiento. Este algoritmo recibe un video de entrada (V) compuesto por N_V fotogramas cada uno de fil_V filas y col_V columnas y lo compara con un video de referencia (T) almacenado en el computador, compuesto por N_P fotogramas cada uno de fil_P filas y col_P columnas, y cuyas dimensiones son menores que las de V . El video de referencia (plantilla T) contiene un movimiento determinado el cual es comparado con el video de entrada para generar un video resultado (R) donde se señala el segmento de V más similar al movimiento descrito en T . Específicamente, el video resultado R es el mismo video V ; pero iluminado en diferentes colores indicando la similitud del movimiento respecto a T . Además, el sistema genera un archivo con los datos espacio-temporales de todo el proceso, incluyendo los datos del instante en que se presentó el movimiento más similar al de la plantilla T .

3.1 Equipo empleado

El sistema se implementó en un Computador Personal Toshiba Satellite con un procesador Intel(R) Dual Core de 2.00 GHz, memoria RAM de 2.00 GB y disco duro de 160GB con Sistema Operativo Windows Vista. El algoritmo se desarrolló en Microsoft Visual C++ utilizando la librería OpenCV para el procesamiento de imágenes y video (Ver Anexo 1).

3.2 Condiciones de los videos

Tanto la plantilla como los videos a analizar deben cumplir con unas especificaciones básicas para ser correctamente procesados por el sistema. Entre estas especificaciones se tiene:

- Formato de video: Los archivos de video deben estar en formato .AVI
- Número de fotogramas: La plantilla puede tener un máximo de 50 fotogramas. Aunque el video a analizar no tiene un límite de fotogramas definido, dependiendo de la capacidad de procesamiento del equipo es posible que si el video es muy largo el sistema no pueda procesarlo.

- Origen del video: Los videos deben ser naturales y no sintéticos. Pruebas realizadas con videos sintéticos (hechos de imágenes por computadora y no de imágenes reales) produjeron resultados desfavorables debido a la textura plana de las imágenes.
- Fondo: El fondo de los videos debe ser estático y es deseable que tenga poco ruido.

3.3 Clasificación de los videos

Los videos se almacenan en el Disco Duro y son leídos y procesados por medio de funciones específicas de la librería OpenCV [7] para el procesamiento de imágenes y video. Estos videos, se clasifican en 2 clases: plantilla (T) y video a analizar (V).

La plantilla T es un video de consulta que contiene el movimiento específico que se desea encontrar en el video a analizar V .

3.4 Diagrama en bloques del sistema

Como se mencionó anteriormente, el sistema desarrollado tiene como objetivo detectar un movimiento contenido en una plantilla T , en un video de entrada V (que es igual o más largo que el video de consulta T). Para desarrollar el algoritmo propuesto, el sistema se dividió en varios bloques, como se muestra en la Figura 1.

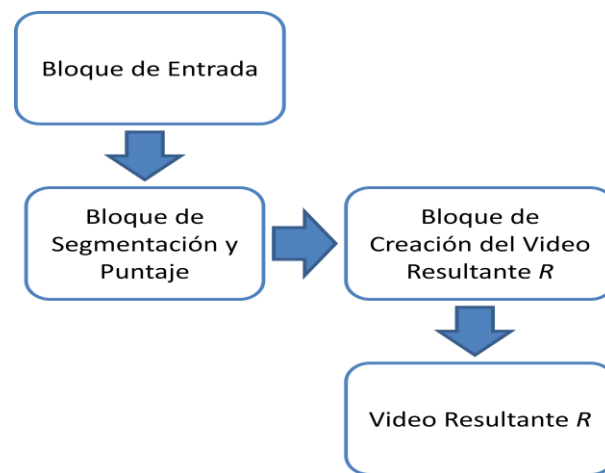


Figura 1. Diagrama en bloques del sistema

El primer bloque es considerado como el bloque de entrada y es quien recibe las 3 entradas del sistema y se encarga de procesarlas de la manera adecuada. El sistema recibe 3 entradas que se detallan brevemente a continuación.

La entrada 1 (plantilla espacio-temporal T) es una pequeña secuencia de video compuesta por N_T fotogramas cada uno de fil_T filas y col_T columnas de menor tamaño respecto al video de entrada que se quiere analizar, tanto espacial como temporalmente, y que contiene el patrón de movimiento a detectar. Esta plantilla debe contener un movimiento bien definido y una cantidad adecuada de fotogramas, dependiendo del movimiento que se quiera analizar. La plantilla puede ser obtenida directamente del video original haciendo un previo recorte del mismo, o puede provenir de un video diferente. En la Figura 2 se observan dos ejemplos de plantillas.



Figura 2. Ejemplo de dos plantillas T: a. plantilla de una persona caminando y **b.** plantilla de un giro de ballet.

La entrada 2 (video a analizar V) es la entrada principal del sistema y está compuesto por N_V fotogramas cada uno de fil_V filas y col_V columnas igual o mayor al tamaño de la plantilla T . Este video se va a descomponer de manera espacio-temporal en segmentos de tamaño igual al de la plantilla; es decir, si se tiene una plantilla T de tamaño $(m \times n \times p)$ se tomarán porciones del video V del mismo tamaño, que serán comparados con respecto a la plantilla. En la Figura 3 se observan algunos fotogramas de dos ejemplos de videos.

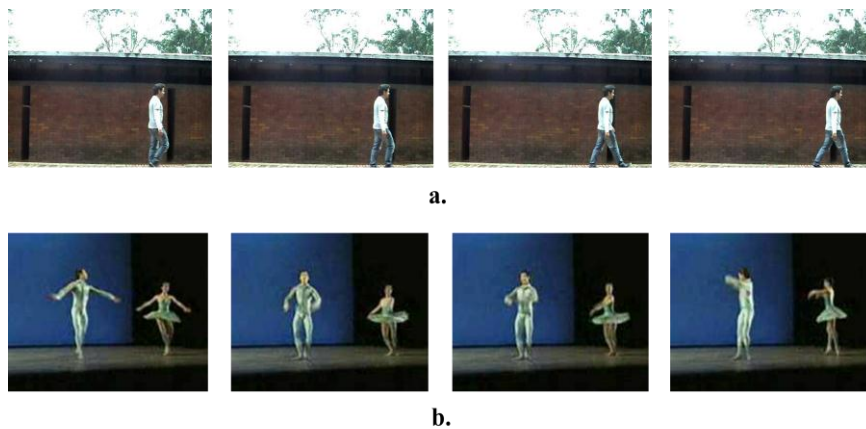


Figura 3. Algunos fotogramas de video de entrada: a. persona caminando y **b.** presentación de la compañía Birmingham Royal Ballet del Reino Unido.

La entrada 3 (especificaciones del análisis ingresadas por el usuario) son algunos parámetros que ingresa el usuario, de acuerdo a la manera en que quiere realizar el análisis del movimiento. Entre estos parámetros, se encuentran el factor de diezmado, el desplazamiento espacial, y el desplazamiento temporal. Cada uno de estos se explican en la Sección 3.

El Bloque de segmentación y puntaje de correlación es el encargado de segmentar el video V en sub-plantillas S del tamaño de la plantilla T y realizar el procesamiento necesario para encontrar el mayor puntaje de correlación entre todas las sub-plantillas S .

Por último se encuentra el Bloque de creación del video resultado R , que se encarga de visualizar sobre el video V la detección del movimiento más similar a la plantilla T .

En la Figura 4 se observa un diagrama en bloques más detallado del sistema. Cada uno de los bloques señalados allí, se explican en la Sección 3.

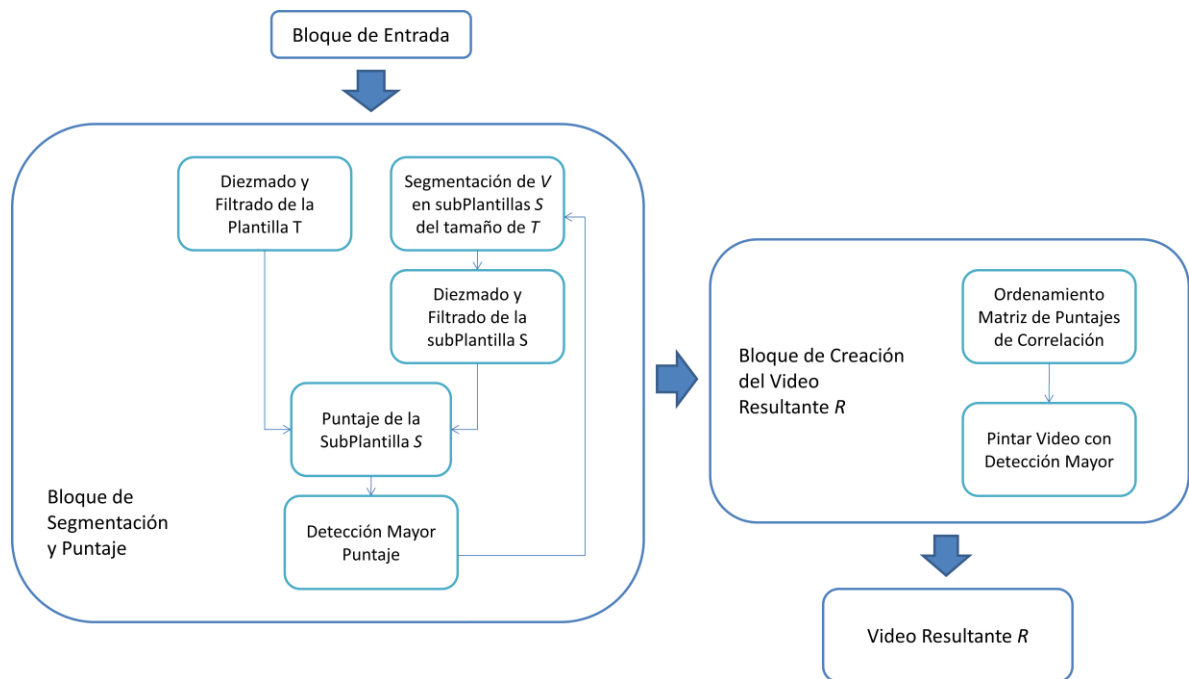
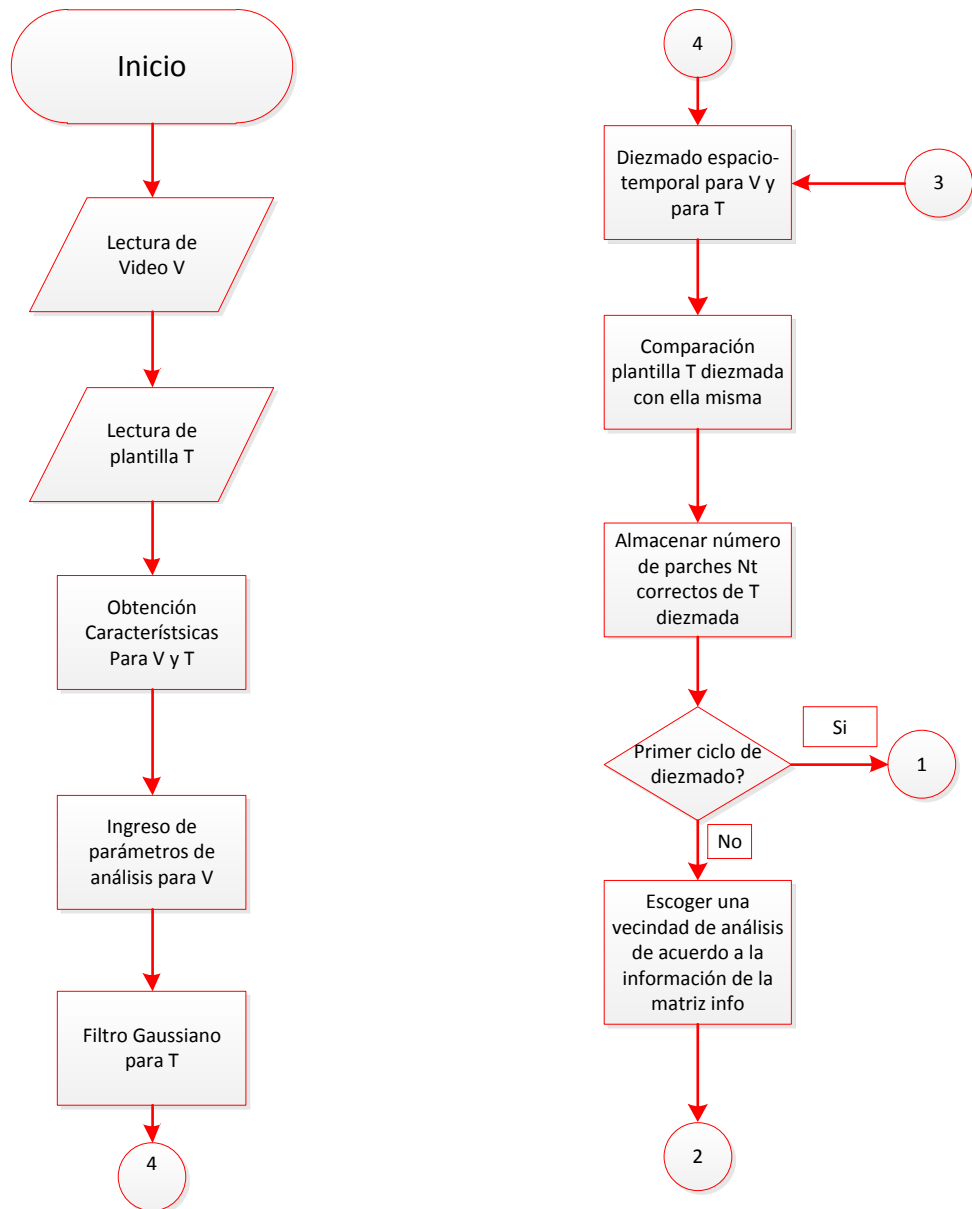


Figura 4. Diagrama en bloques del sistema (en detalle).

3.5 Diagrama de flujo

En la Figura 5 se presenta el diagrama de flujo del sistema de detección de patrones de movimiento.



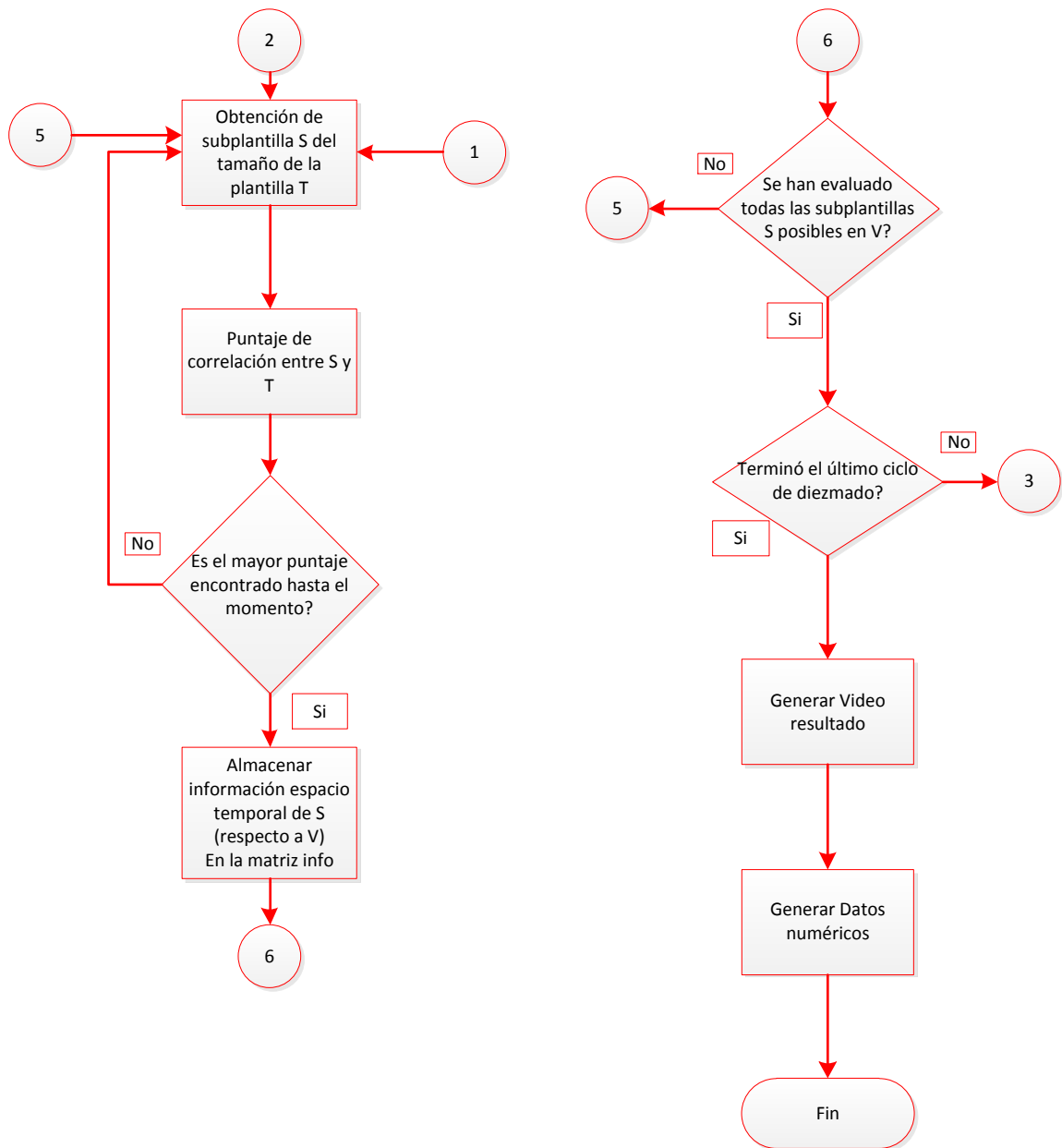


Figura 5. Diagrama de flujo principal.

En la Figura 6 se presenta el diagrama de flujo para el Bloque de detección puntaje de correlación entre S y T .

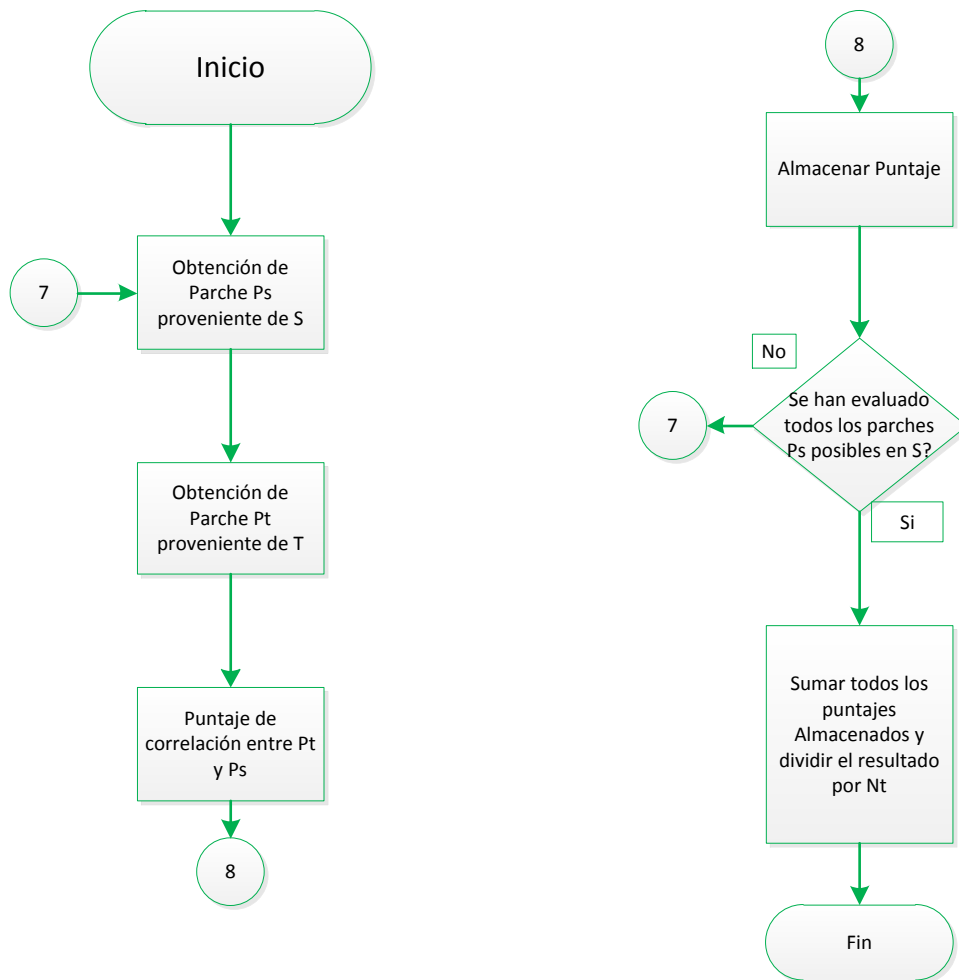


Figura 6. Diagrama de flujo del bloque puntaje de correlación entre S y T .

4. DESARROLLO TEÓRICO

El algoritmo desarrollado está basado en el método propuesto por Eli Shechtman y Michal Irani en su artículo titulado *Space-Time Behavior-Based Correlation* [5] que fue estudiado en detalle, y posteriormente se implementó una versión del algoritmo en C++ propia de los autores de este trabajo de grado. A saber, algunas variaciones al algoritmo original fueron desarrolladas así como la estrategia de visualización de los resultados.

Como se ha mencionado anteriormente, el sistema desarrollado permite introducir una secuencia del video de interés V compuesto por N_V fotogramas cada uno de fil_V filas y col_V columnas para “correlacionarlo” con un pequeño segmento de video conocido como plantilla espacio-temporal T compuesto por N_T fotogramas cada uno de fil_T filas y col_T columnas que contiene un movimiento de interés. Para esto, el video V es dividido en segmentos espacio-temporales S (llamados sub-plantillas) del tamaño de la plantilla T , cada uno de los cuales es correlacionado con la misma T . A partir de este procedimiento se obtiene una serie de puntajes de correlación en diferentes posiciones espacio-temporales, creando una superficie de correlación que se verá reflejada en el video resultado R . Esta superficie de correlación se obtiene computando las medidas de correlación local de pequeños segmentos espacio-temporales que se llamarán parches (por ejemplo, compuestos por 3 fotogramas cada uno de 7 filas y 7 columnas) y que están contenidos en S y T . A saber, cada sub-plantilla S se divide en pequeños parches espacio-temporales P_S ubicados en cada punto (i,j,k) de S , y se comparan con sus correspondientes parches espacio-temporales P_T de la plantilla T (también ubicados en los mismos puntos (i,j,k) pero de T). Los puntajes locales resultantes de cada correlación entre parches correspondientes, se suman hasta obtener un puntaje global de correlación para esta sub-plantilla S con respecto a la plantilla T . Este puntaje global obtenido, indica el grado de similitud entre los dos movimientos descritos por las secuencias de video S y T . La Figura 7 muestra una explicación gráfica de cada uno de los elementos mencionados recientemente.

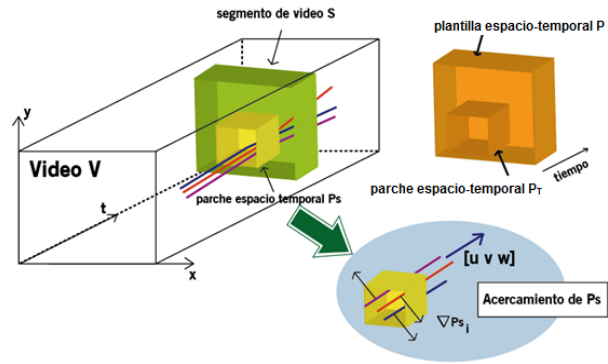


Figura 7. Notaciones y componentes del sistema.

4.1 Segmentación y puntaje de correlación

Una vez las respectivas entradas son ingresadas al sistema, éste se encarga de procesarlas adecuadamente de acuerdo a las funciones programadas del Bloque de segmentación y puntaje de correlación. En la Figura 8 se observa un diagrama de cada uno de los componentes de este Bloque.

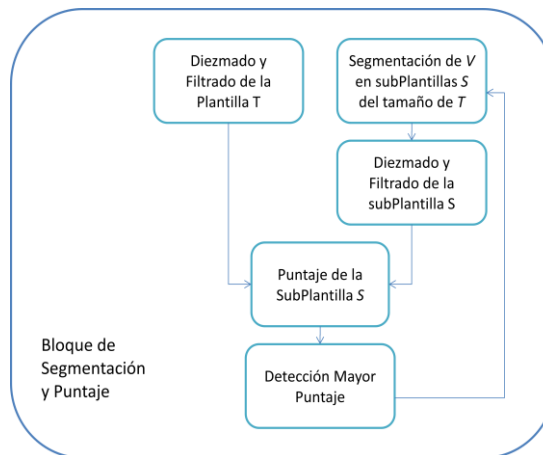


Figura 8. Bloque de segmentación y puntaje de correlación.

Este Bloque está compuesto de varias funciones específicas que se explican a continuación.

4.1.1 Obtención de sub-plantillas de V

De acuerdo al algoritmo descrito en [5], la comparación de movimientos debe ser entre secuencias de video de iguales dimensiones espacio-tiempo. Es por esto que el video a analizar V es segmentado en secuencias de video de las mismas dimensiones que la plantilla T . Para esto se toman segmentos en diferentes posiciones espacio-temporales de V denominados sub-plantillas S ,

y los cuales son analizados bajo el procesamiento normal del algoritmo para determinar la similitud respecto a la plantilla T . Una vez comparados todos los segmentos del video V se detecta la ubicación de movimiento más similar al descrito por la plantilla T y se procede a la creación del video resultado R .

La Figura 9 muestra la manera en que se recorre el video V para seleccionar las sub-plantillas S . El desplazamiento espacio-temporal entre sub-plantillas es determinado por el usuario en la entrada 3, teniendo en cuenta que entre menor sea el desplazamiento, menor es la resolución espacial y temporal del análisis, y por tanto se puede perder mucha información.

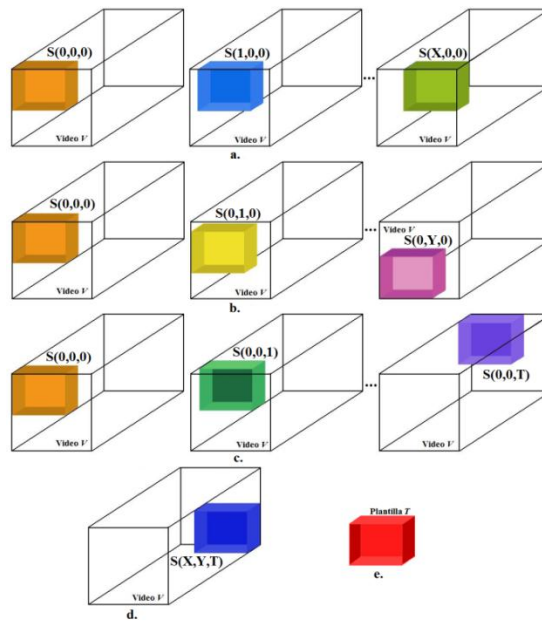


Figura 9. Recorrido y segmentación por sub-plantillas S de todo el video V . a. Recorrido espacial a lo largo del eje x. b. Recorrido espacial a lo largo del eje y. c. Recorrido temporal a lo largo del eje t. d. Última sub-plantilla después de recorrer todo el video V . e. plantilla T , que es de igual tamaño que las sub-plantillas S .

4.1.2 Diezmado y filtrado de la plantilla T y la sub-plantilla S (pre-procesamiento)

Si se computa el volumen total de correlación de una secuencia de video como las utilizadas en las pruebas (por ejemplo el video de la compañía Birmingham Royal Ballet del Reino Unido, 144 pixeles x 192 pixeles x 766 fotogramas) con una plantilla también utilizada en las pruebas (por ejemplo, 101 pixeles x 56 pixeles x 22 fotogramas), es decir, si se realiza el procesamiento con el mínimo desplazamiento (a saber, pixel a pixel), el tiempo de procesamiento en un computador con procesador Intel(R) Dual Core de 2.00 GHz y con Sistema Operativo Windows Vista sería

extremadamente alto y la memoria virtual del equipo no sería suficiente para procesar y almacenar tanta información.

Sin embargo, como sólo es de interés el pico más alto de correlación, no es necesario analizar todo el video pixel a pixel y entonces es posible realizar una búsqueda multiresolución, empezando desde el video con baja resolución hasta el video con mayor resolución. De esta manera el tiempo de procesamiento se reduce sustancialmente.

Para tal fin se realiza un diezmado espacio-temporal del video a analizar V y de la plantilla T en un factor X_1 (a saber, se seleccionan las dos secuencias de video y se baja la resolución en un valor de 2^{X_1}) y se realiza una búsqueda del mayor pico de correlación entre plantillas y sub-plantillas a lo largo de todo el video. Una vez se tiene este pico, se toman sus coordenadas y se realiza una búsqueda en un nivel de resolución mayor en espacio y tiempo de ambas secuencias T y V (factor de diezmado X_2) en una pequeña vecindad alrededor de este pico para encontrar las posiciones más exactas. Este proceso se realiza sucesivamente para los niveles de resolución mayores hasta que la búsqueda final en la secuencia de resolución más alta (video y plantilla originales) determina la ubicación exacta del pico más alto de correlación. Además, para evitar efectos negativos en las comparaciones a causa de movimientos muy rápidos, se aplicó un filtro espacial Gaussiano tanto para la sub-plantilla S como para la plantilla T , logrando un desenfoque apropiado de ambas secuencias.

4.1.3 Puntaje de correlación de la sub-plantilla S

Una vez segmentado el video largo V y seleccionada una sub-plantilla S del tamaño de la plantilla T , se procede a realizar la comparación entre los movimientos contenidos en cada segmento de video (S y T) para obtener un puntaje de correlación global. A saber, se realiza un análisis de correlación espacio-temporal entre la plantilla T y la sub-plantilla S , que consta de varios bloques que se muestran en la Figura 10.

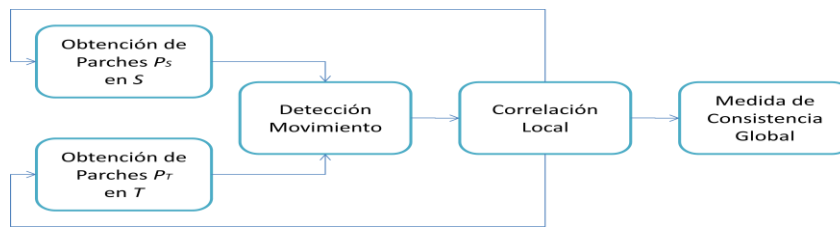


Figura 10. Diagrama en bloques que describe el proceso para encontrar el puntaje de correlación de las sub-plantillas S .

4.1.3.1 Obtención de parches para T y S

Para realizar la comparación entre los dos segmentos de video S y T , en primer lugar se debe segmentar cada uno de ellos en fracciones más pequeñas que llamaremos parches espacio-temporales. Los parches espacio-temporales son las secuencias de video más pequeñas a procesar en este sistema y resultan de la partición de la plantilla T y de cada sub-plantilla S . Además, para la versión del algoritmo desarrollada en este proyecto, los parches espacio-temporales tendrán una dimensión temporal de 3 fotogramas para así estandarizar el proceso y hacerlo más sencillo. Las dimensiones espaciales de estos parches son mucho menores respecto a las de la plantilla (en la mayoría de los experimentos se usaron parches de 7×7 píxeles). En la Figura 11 se observan algunos parches de dos plantillas diferentes utilizadas en las pruebas realizadas.

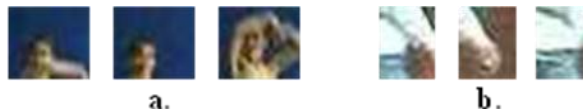


Figura 11. Parches espacio-temporales de dos plantillas de ejemplo compuestos por 3 fotogramas cada uno de 7 píxeles de alto y 7 píxeles de ancho. **a.** Parche espacio-temporal de la plantilla de un giro de ballet. **b.** Parche espacio-temporal de la plantilla de una persona caminando.

La medida de similitud entre estos pequeños parches captura la similitud de los movimientos subyacentes implícitos. Para un parche espacio-temporal P suficientemente pequeño (por ejemplo, $7 \times 7 \times 3$), se puede suponer que todos los píxeles dentro de éste se mueven con un mismo movimiento. En secuencias de video naturales este supuesto casi siempre es verdadero pues no hay cambios abruptos entre uno y otro fotograma, es decir que muy pocos parches en una plantilla y sub-plantilla violarán este supuesto; a saber, sólo aquellos que se encuentran localizados en discontinuidades de movimiento, así como los que tienen cambios muy fuertes en la dirección y/o velocidad del movimiento.

Para obtener los diferentes parches espacio-temporales se realiza una división de T y S en pequeños segmentos espacio-temporales que se notan como P_{Ti} y P_{Si} respectivamente, donde ($i = 1, 2, \dots, np$) y np es el número de píxeles dentro del parche (por ejemplo, si las dimensiones espacio-temporales del parche son $7 \times 7 \times 3$, entonces el número de píxeles dentro de este parche es $np = 147$). Estos parches se correlacionan entre ellos (parches correspondientes de T y S) y se comparan con el fin de encontrar un puntaje local de correlación entre ellos. De manera similar a la forma en que se recorre V a medida que se obtienen diferentes sub-plantillas en V , se lleva a cabo el recorrido y la obtención de los parches espacio-temporales en S (P_S) y T (P_T) respectivamente. Al finalizar el análisis de dos parches, se selecciona otro par de parches P_T y P_S y se procede a realizar de nuevo el procesamiento por parte del sistema. Para lograr una aceleración en el procesamiento por parte del sistema, se puede seleccionar sólo un conjunto de parches que represente suficientemente bien la plantilla T , sin necesidad de tomar parches traslapados alrededor de todos los píxeles en la plantilla (y sus parches correspondientes en la sub-plantilla S). Para la versión del algoritmo implementada en este Trabajo de Grado, el conjunto de parches fue seleccionado de tal manera que no hubiese traslape entre ellos, es decir que cada parche está inmediatamente seguido del otro pero sin compartir píxeles entre ellos.

4.1.3.2 *Detección movimiento (modificación a [5])*

Después de segmentada la plantilla T y la sub-plantilla S en parches espacio-temporales P_T y P_S , se debe proceder con la correlación espacio-temporal entre ellos para obtener un puntaje que indique la similitud entre el movimiento implícito en estos. Sin embargo, antes de realizar esta operación que se explica en detalle en la Sección 3.1.3.3, se implementó un bloque adicional (no considerado en el algoritmo original [5]) que se encarga de despreciar aquellos parches que no poseen movimiento alguno, es decir, que la intensidad en cada uno de sus píxeles no cambia durante los 3 fotogramas de duración del parche. Todo esto teniendo en cuenta algunos resultados desfavorables obtenidos durante la etapa de desarrollo del algoritmo en los que se encontró que el sistema detectaba altos puntajes de correlación en parches uniformes (que pertenecían al fondo). Sean $P_1(x, y, t_1)$, $P_2(x, y, t_2)$, $P_3(x, y, t_3)$, ..., $P_n(x, y, t_n)$ píxeles en un parche espacio-temporal en n instantes de tiempo diferentes (en nuestra versión del algoritmo, n es igual a 3), pero en la misma posición espacial. Se puede hallar el promedio de la intensidad de estos píxeles a través de la siguiente expresión:

$$\overline{\mu_p} = \frac{1}{n} \sum_{t=1}^n P_n(x, y, t) \quad (1)$$

Ahora se puede obtener una “medida” de la cantidad de variación a través del tiempo entre estos pixeles de la siguiente manera:

$$\sigma_{pxy} = \frac{1}{n} \sum_{t=1}^n (P_n(x, y, t) - \overline{\mu_p})^2 \quad (2)$$

Sin embargo, como esta medida temporal se calcula para todos los pixeles de un parche, si se cuenta con un parche espacio-temporal de tamaño (r filas x c columnas x n fotogramas), se obtienen un conjunto de rc medidas; por lo que es necesario tratar de obtener una medida general que caracterice la variación temporal de los pixeles del parche.

Sea $\overline{\sigma_{prc}}$ el promedio de todas estas rc medidas obtenidas anteriormente; luego la medida de desviación completa está determinada por la resta entre cada medida obtenida y el promedio de estas, de la siguiente manera:

$$d = \frac{1}{rc} \sum_{p=1}^N (\sigma_{prc} - \overline{\sigma_{prc}})^2 \quad (3)$$

Si d es un valor muy cercano a cero, es evidente que el movimiento en ese parche durante ese número de fotogramas establecido (en nuestro caso 3) es mínimo, y por lo tanto puede descartarse.

4.1.3.3 Correlación Local

Una vez descartados aquellos parches uniformes y/u obtenidos dos parches espacio-temporales P_T y P_S (uno de la plantilla T y otro de la sub-plantilla S) no uniformes, continuamos con el análisis de correlación espacio-tiempo para determinar una medida de consistencia entre estos. En la Figura 12 se observan los bloques que componen el proceso para encontrar esta medida.

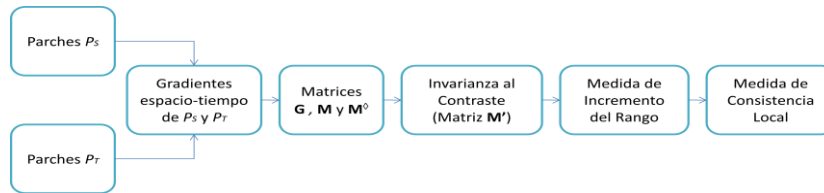


Figura 12. Diagrama en bloques del proceso de correlación local.

4.1.3.3.1 Gradientes espacio-tiempo de P_S y P_T

Cada movimiento que sea uniforme dentro de un parche espacio-temporal, induce una serie de líneas rectas de intensidad dentro de éste, que apuntan en una misma dirección espacio-tiempo (u, v, w) como se observa en la Figura 13.

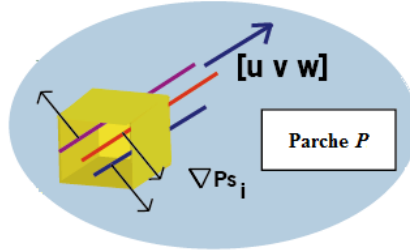


Figura 13. Líneas de intensidad inducidas por un movimiento uniforme dentro de un parche P .

Esta dirección (u, v, w) depende de la dirección del movimiento y por esto generalmente es diferente para cada punto (i, j, k) en la secuencia de video y por tanto para cada parche espacio-temporal; es decir, se asume que el movimiento es uniforme sólo localmente (dentro del parche).

Los gradientes espacio-tiempo de la intensidad en cada pixel dentro de los parches de T y S (P_T y P_S respectivamente) son determinados por la ecuación (4); donde $(i = 1, 2, \dots, np)$ y np es el número de pixeles dentro del parche.

$$\nabla P_i = (P_{xi}, P_{yi}, P_{ti}) \quad (4)$$

Por tanto, estos gradientes estarán orientados hacia la dirección de máximo cambio de la intensidad en espacio y tiempo, es decir, que estos gradientes serán perpendiculares a la dirección (u, v, w) de las líneas de intensidad inducidas por el movimiento. De esto se deduce la ecuación (5).

$$\nabla P \cdot \begin{bmatrix} u \\ v \\ w \end{bmatrix} = 0 \quad (5)$$

Es importante aclarar que no necesariamente los gradientes de todos los pixeles están orientados en la misma dirección (o son paralelos entre sí), pero todos sí están ubicados en un mismo plano 2D normal a (u, v, w) .

4.1.3.3.2 Matrices G , M y M^\diamond

Una vez obtenidos los gradientes para todos los np pixeles en el parche, se pueden reunir en una sola matriz que se llamará Matriz Gradiente \mathbf{G} , y así obtener:

$$G = \begin{bmatrix} Px_1 & Py_1 & Pt_1 \\ Px_2 & Py_2 & Pt_2 \\ \dots & \dots & \dots \\ Px_{np} & Py_{np} & Pt_{np} \end{bmatrix}_{np \times 3} \quad (6)$$

Por tanto, de la ecuación (5) y la ecuación (6) se tiene que:

$$\begin{bmatrix} Px_1 & Py_1 & Pt_1 \\ Px_2 & Py_2 & Pt_2 \\ \dots & \dots & \dots \\ Px_n & Py_n & Pt_n \end{bmatrix}_{nx3} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \dots \\ 0 \end{bmatrix}_{nx1} \quad (7)$$

Multiplicando ambos lados de la ecuación (7) por \mathbf{G}^T (la transpuesta de la Matriz Gradiente \mathbf{G}), se obtiene:

$$G^T G \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}_{3 \times 1} \quad (8)$$

$\mathbf{G}^T \mathbf{G}$ es una matriz de 3×3 , que se llamará “Matriz Gram” de \mathbf{G} y se denominará \mathbf{M} . El tamaño de esta matriz es independiente del tamaño del parche espacio-temporal:

$$M = G^T G = \begin{bmatrix} \sum P_x^2 & \sum P_x P_y & \sum P_x P_t \\ \sum P_y P_x & \sum P_y^2 & \sum P_y P_t \\ \sum P_t P_x & \sum P_t P_y & \sum P_t^2 \end{bmatrix}_{3 \times 3} \quad (9)$$

Las sumatorias de la ecuación (9) son para todos los pixeles dentro de los parches espacio-temporales.

Por tanto, para todos los parches que contienen un **único** movimiento uniforme, La matriz $\mathbf{M}_{3 \times 3}$ es una matriz de rango-deficiente, es decir, $\text{rango}(\mathbf{M}) \leq 2$. Esto implica que su más pequeño valor-propio es cero ($\lambda_{\min} = 0$) y su correspondiente vector-propio es (u, v, w) .

Por otro lado, si existe un parche espacio-temporal que no contenga un **único** movimiento uniforme (es decir, que hay más de un vector $[u \ v \ w]$ que es perpendicular al máximo cambio de intensidad en espacio y tiempo), entonces $\text{rango}(\mathbf{M}) = 3$; a saber, este parche espacio-temporal fue inducido por varios movimientos independientes. Esto ocurre cuando el parche está localizado en discontinuidades de movimiento. Sin embargo, en secuencias naturales de video estos parches son atípicos.

Como lo que se quiere es detectar si dos parches espacio-temporales (específicamente P_T y P_S) con patrones de intensidad completamente distintos, pueden haber resultado de un vector de movimiento similar, se puede extender el análisis del rango de una matriz \mathbf{M} como se explicará a continuación.

Se dice que dos parches espacio-temporales (P_T y P_S) **son consistentes en su movimiento** si existe un vector común $\mathbf{u} = [u, v, w]^T$ que cumpla la ecuación (7) para ambos parches, es decir $\mathbf{G}_T \mathbf{u} = \mathbf{0}$ y $\mathbf{G}_S \mathbf{u} = \mathbf{0}$. Al concatenar ambas Matrices se tiene:

$$\mathbf{G}_{TS} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} \mathbf{G}_T \\ \mathbf{G}_S \end{bmatrix}_{2n \times 3} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ \cdot \\ \cdot \\ 0 \end{bmatrix}_{2n \times 1} \quad (10)$$

Donde \mathbf{G}_{TS} contiene todos los gradientes de intensidad espacio-tiempo de ambos parches P_T y P_S . Luego, si se multiplican ambos lados de la ecuación (10) por \mathbf{G}_{TS}^T , se obtiene:

$$\mathbf{M}_{TS} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}_{3 \times 1} \quad (11)$$

Donde $\mathbf{M}_{TS} = \mathbf{G}_{TS}^T \mathbf{G}_{TS}$ es una matriz de rango-deficiente, es decir, $\text{rango}(\mathbf{M}_{TS}) \leq 2$.

Por otro lado, se dice que dos parches espacio-temporales (P_T y P_S) **NO son consistentes en su movimiento** si la matriz combinada \mathbf{M}_{TS} NO es de rango-deficiente, esto es, $\text{rango}(\mathbf{M}_{TS}) = 3$; y

por lo tanto su más pequeño valor-propio NO es cero $\lambda \min(M_{12}) \neq 0$. Esto implica que cada parche fue inducido por un único y diferente movimiento uniforme.

Se puede apreciar que $\mathbf{M}_{TS} = \mathbf{M}_T + \mathbf{M}_S = \mathbf{G}_T^T \mathbf{G}_T + \mathbf{G}_S^T \mathbf{G}_S$.

Sin embargo, la restricción de $\text{rango}(\mathbf{M}_{TS}) = 3$ para detectar inconsistencias de movimiento en dos parches es una condición suficiente pero no necesaria; es decir que si esta condición se cumple, entonces, por un lado implica que NO existe un único movimiento que pueda inducir el patrón de intensidad de ambos parches espacio-temporales (P_T y P_S) y por lo tanto estos NO son consistentes en el movimiento, pero por otro lado, pueden haber casos en donde NO existe un único movimiento uniforme que induzca los patrones de intensidad espacio-tiempo de los dos parches, y aun así $\text{rango}(\mathbf{M}_{TS}) < 3$.

El único caso en el cual la restricción de $\text{rango}(\mathbf{M}_{TS}) = 3$ es tanto suficiente como necesaria para detectar inconsistencias en el movimiento, es cuando ambas matrices \mathbf{M}_T y \mathbf{M}_S son cada una de rango-2, a saber, cuando ambos parches espacio-temporales P_T y P_S contienen características degeneradas de la imagen (corner-like).

4.1.3.3.3 Invarianza al contraste (Matriz \mathbf{M}')

En el próximo bloque se explicará cómo obtener una condición necesaria y suficiente para detectar la consistencia o inconsistencia entre los movimientos de dos parches; sin embargo, por ahora se presentará una mejora a la matriz \mathbf{M} que la hace invariante al contraste de los parches espacio-temporales.

Si por ejemplo se tiene que el parche espacio-temporal de la plantilla T (P_T) tiene un contraste mucho mayor al del parche espacio-temporal de la sub-plantilla S (P_S); esto, matemáticamente, implica que los gradientes espacio-tiempo de P_T son mucho más grandes que los de P_S , y por lo tanto que la matriz \mathbf{M}_T va a tener un mayor efecto en la matriz \mathbf{M}_{TS} que el efecto que tendrá \mathbf{M}_S .

Por esto, para que la medida de consistencia sea capaz de comparar parches espacio-temporales de contrastes muy diferentes, se procede a normalizar los gradientes del parche en cada pixel por la magnitud media de los gradientes locales de éste. Por tanto, se reemplaza la matriz $\mathbf{M} = \mathbf{G}^T \mathbf{G}$ por la nueva matriz $\mathbf{M}' = \mathbf{G}^T \mathbf{W}^{-1} \mathbf{G}$, donde \mathbf{W} es una matriz diagonal con las magnitudes medias de los gradientes locales en su diagonal.

Entonces, para un parche espacio-temporal P , se tiene:

$$W = \begin{bmatrix} \bar{\nabla}P_1 & 0 & \dots & 0 \\ 0 & \bar{\nabla}P_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \bar{\nabla}P_{np} \end{bmatrix}_{np \times np} \quad (12)$$

Donde: $\bar{\nabla}P_1 = \sqrt{P_{x1}^2 + P_{y1}^2 + P_{t1}^2}$, $\bar{\nabla}P_2 = \sqrt{P_{x2}^2 + P_{y2}^2 + P_{t2}^2}$, \dots , $\bar{\nabla}P_n = \sqrt{P_{xn}^2 + P_{yn}^2 + P_{tn}^2}$

4.1.3.3.4 Medida de Incremento del Rango

Ahora, volviendo a la búsqueda de una restricción de consistencia de movimiento suficiente y necesaria para todo tipo de parches espacio-temporales, se procede a examinar todos los posibles rangos de la matriz \mathbf{M}' de un parche espacio-temporal P que contiene un único movimiento uniforme. De esta manera, se tienen los siguientes casos:

- $\text{rango}(\mathbf{M}') = 2$ cuando P contiene una imagen característica como esquinas (corner-like).
- $\text{rango}(\mathbf{M}') = 1$ cuando P contiene una imagen característica como bordes (edge-like).
- $\text{rango}(\mathbf{M}') = 0$ cuando P contiene una imagen de una región uniformemente coloreada.

Toda esta información (sobre las propiedades espaciales de P) está contenida en la siguiente matriz de 2×2 M^\diamond , que es la matriz secundaria (superior-izquierda) de \mathbf{M}' .

$$M^\diamond = \begin{bmatrix} \sum P_x^2 & \sum P_x P_y \\ \sum P_y P_x & \sum P_y^2 \end{bmatrix}_{2 \times 2} \quad (13)$$

Cuando se tiene un parche con un único movimiento uniforme, el componente temporal sumado en la matriz \mathbf{M}' (tercera columna y tercera fila) no introduce ningún incremento en el rango y por lo tanto se tiene la siguiente condición: $\text{rango}(\mathbf{M}') = \text{rango}(M^\diamond)$.

Sin embargo, cuando un parche espacio-temporal contiene más de un movimiento (cuando el movimiento no es a lo largo de una única línea recta), la condición de rango anterior no se mantiene. En esos casos, la componente temporal que se suma en la matriz \mathbf{M}' produce un incremento en el rango, a saber, $\text{rango}(\mathbf{M}') = \text{rango}(M^\diamond) + 1$. Debido a que al pasar de M^\diamond a \mathbf{M}' sólo se suma una fila y una columna, la diferencia en el rango no puede ser mayor a 1.

De acuerdo a lo último, para determinar cuándo un parche posee un único o múltiples movimientos, se puede usar la siguiente medida de incremento del rango:

$$\Delta r = \text{rango}(M) - \text{rango}(M^\diamond) \quad (14)$$

Si $\Delta r = 0$, implica que el parche posee un único movimiento; pero si $\Delta r = 1$, implica que ese parche contiene varios movimientos.

Si se extiende el análisis anterior para dos diferentes parches espacio-temporales, se obtiene la siguiente condición (que es necesaria y suficiente para detectar si dos parches son consistentes o inconsistentes en su movimiento):

$$\Delta r_{TS} = \text{rango}(M'_{TS}) - \text{rango}(M_{TS}^\diamond) \quad (15)$$

Si $\Delta r = 0$, implica que los dos parches son consistentes en su movimiento (poseen el mismo movimiento); pero si $\Delta r = 1$, implica que los dos parches son inconsistentes en su movimiento (contienen dos movimientos distintos).

Sin embargo, esta medida de consistencia presenta un grave problema para lo que se pretende en este proyecto. Si un movimiento es muy similar pero no idéntico al de la plantilla, la medida de incremento del rango (que se tiene en este momento) indicaría que los movimientos no son consistentes, pues solamente arroja valores binarios (0 ó 1). Además, como el rango de una matriz es determinado por el número de valores-propios diferentes de cero que ésta tenga y cómo en el caso de un video natural, a causa del ruido, los valores-propios nunca son cero; cuando el sistema procese videos que presenten movimientos muy similares al de la plantilla (no exactamente iguales) y que tengan algo de ruido (como es de esperarse), con seguridad los resultados obtenidos de la comparación con la plantilla T serán completamente erróneos.

Por lo tanto, se necesita una medida de consistencia de movimiento entre dos parches espacio-temporales, que permita obtener valores continuos entre 0 y 1 del incremento del rango de la matriz \mathbf{M}' . La obtención de esta medida continua, se detalla a continuación:

Sean $\lambda_1 \geq \lambda_2 \geq \lambda_3$ los valores-propios de la matriz de 3×3 \mathbf{M}' . Sean $\lambda_1^\diamond \geq \lambda_2^\diamond$ los valores-propios de su matriz secundaria (superior-izquierda) de 2×2 \mathbf{M}^\diamond . De la Propiedad de Entrelazar de los valores-propios en las matrices simétricas [8] (explicada en la Sección 1), se tiene que

$\lambda_1 \geq \lambda_1^\diamond \geq \lambda_2 \geq \lambda_2^\diamond \geq \lambda_3$. De lo anterior y de las propiedades de los valores-propios de una Matriz, se obtienen las siguientes ecuaciones:

$$\lambda_1 \geq \frac{\lambda_1 \cdot \lambda_2 \cdot \lambda_3}{\lambda_1^\diamond \cdot \lambda_2^\diamond} = \frac{\det(\mathbf{M}')}{\det(\mathbf{M}^\diamond)} \geq \lambda_3 \quad (16)$$

y

$$1 \geq \frac{\lambda_2 \cdot \lambda_3}{\lambda_1^\diamond \cdot \lambda_2^\diamond} \geq \frac{\lambda_3}{\lambda_1} \geq 0 \quad (17)$$

De la ecuación (17) se define la medida continua de incremento del rango (que está entre 0 y 1) como:

$$\Delta r = \frac{\lambda_2 \cdot \lambda_3}{\lambda_1^\diamond \cdot \lambda_2^\diamond} \quad (18)$$

La nueva definición de incremento del rango, descrita por la ecuación (18), permite manejar videos ruidosos y obtener una medida proporcional al grado de la consistencia en los movimientos.

Sin embargo, aunque esta medida es bastante acertada, la estimación de los valores-propios de la matriz \mathbf{M}_{TS} en cada pixel es bastante costosa computacionalmente; y por esto, es adecuado implementar una nueva medida que permita obtener una aproximación del incremento del rango, sin tener que computar los valores-propios.

Entonces, como $\det(\mathbf{M}') = \lambda_1 \cdot \lambda_2 \cdot \lambda_3$, y $\det(\mathbf{M}^\diamond) = \lambda_1^\diamond \cdot \lambda_2^\diamond$; se tiene de la ecuación (18) que:

$$\Delta r = \frac{\lambda_2 \cdot \lambda_3}{\lambda_1^\diamond \cdot \lambda_2^\diamond} = \frac{\det(\mathbf{M}')}{\det(\mathbf{M}^\diamond) \cdot \lambda_1} \quad (19)$$

Si se tiene que la Norma de Frobenius de la matriz \mathbf{M}' es $\|\mathbf{M}'\|_F = \sqrt{(\sum (i, j)^2)}$, entonces se sostiene la siguiente relación entre $\|\mathbf{M}'\|_F$ y λ_1 :

$$\lambda_1 \leq \|\mathbf{M}'\|_F \leq \sqrt{3}\lambda_1 \quad (20)$$

Donde $\sqrt{3} (\approx 1.73)$ está relacionado con la dimensión de \mathbf{M}' (3 x 3). Por lo tanto, la medida de incremento del rango Δr puede ser aproximada a:

$$\Delta\hat{r} = \frac{\det(M^\circ)}{\det(M^\diamond) \cdot \|M^\circ\|_F} \quad (21)$$

Aunque $\Delta\hat{r}$ es menos preciso que Δr (ecuación (18)), esta nueva medida no requiere computar valores-propios lo cual acelera el proceso de correlación espacio-tiempo y provee la suficiente precisión para obtener resultados adecuados. A saber, la medida de incremento del rango de la ecuación (14) está limitada por: $\Delta\hat{r} \leq \Delta r \leq \sqrt{3}\Delta\hat{r}$.

4.1.3.3.5 Medida de Consistencia Local

Como se mencionó antes, una plantilla espacio-temporal T es un conjunto de parches espacio-temporales, que es correlacionada con un segmento del mismo tamaño proveniente de un Video largo V , y que se llamará sub-plantilla S . Cada segmento S (o sub-plantilla) estará centrado alrededor de cada punto (x,y,t) en el video largo y se comparará su consistencia con la plantilla T . Una buena coincidencia entre ambas secuencias de video, T y S , debería:

- Minimizar el número de coincidencias locales inconsistentes entre dos parches con un único movimiento lineal (esto es, minimizar la medida conjunta de incremento del rango de dos parches, dada por la ecuación (15))
- Y debería minimizar también el número de coincidencias entre parches que contienen múltiples movimientos o discontinuidades de movimiento (esto es, maximizar la medida individual de incremento del rango para los parches no lineales de ambas secuencias, dada por la ecuación (14)).

La siguiente medida permite obtener bastante bien, el grado de inconsistencia local entre un parche espacio-temporal $P_T \in T$ y un parche espacio-temporal $P_S \in S$.

$$m_{TS} = \frac{\Delta\hat{r}_{TS}}{\min(\Delta\hat{r}_T, \Delta\hat{r}_S) + \varepsilon} \quad (22)$$

Donde ε es un número muy pequeño (por ejemplo, $\varepsilon = 10^{-5}$) que se necesita para evitar la división por 0.

La medida dada por la ecuación (22) produce valores bajos cuando P_T y P_S son consistentes en sus movimientos (esto es, contienen movimientos muy similares); en este caso,

$\Delta\hat{r}_{TS} \approx \Delta\hat{r}_T \approx \Delta\hat{r}_S \approx 0$. Por otro lado, la medida también provee valores bajos cuando ambos parches espacio-temporales P_T y P_S están localizados en discontinuidades de movimiento dentro de sus propias secuencias (T y S respectivamente); en este caso $\Delta\hat{r}_{TS} \approx \Delta\hat{r}_T \approx \Delta\hat{r}_S \approx 1$. En todos los demás casos, m_{TS} proveerá valores altos (que implican inconsistencia en el movimiento).

Algunas evaluaciones matemáticas realizadas por nosotros (y por los autores de [5], según mencionan en su artículo) mostraron siempre que para dos parches espacio-temporales P_T y P_S $\Delta\hat{r}_{TS} \geq \min(\Delta\hat{r}_T, \Delta\hat{r}_S)$. Además si se comparan dos parches idénticos (por ejemplo, P_T con P_T), tenemos que $\Delta\hat{r}_{TT} = \Delta\hat{r}_T$.

De lo anterior, se deduce que la medida m_{TS} tiene un límite inferior de 1, y que la medida más baja posible es cuando se compara un parche espacio-temporal con él mismo.

Sin embargo, para la versión del algoritmo implementada en este Trabajo de Grado, lo que se utilizó fue una medida de **consistencia** entre parches espacio-temporales (lo que se tiene con m_{TS} es una medida de **inconsistencia**); y por lo tanto, para conseguirla se hace uso del inverso de la ecuación (22), obteniendo una medida de **consistencia** entre parches, dada por $1/m_{TS}$.

De lo anterior, se deduce que la medida $1/m_{TS}$ tiene un límite superior de 1, y que la medida más alta posible es cuando se compara un parche espacio-temporal con él mismo (a saber, $0 \leq 1/m_{TS} \leq 1$).

4.1.3.4 Medida de Consistencia Global

Como ya se cuenta con una medida local para la consistencia entre un movimiento descrito por dos parches espacio-temporales P_T y P_S (de la plantilla T y la sub-plantilla S respectivamente), lo que sigue es reunir cada una de estas medidas (a saber, la medida de cada uno de los parches en T y S) en una única medida global de consistencia entre un movimiento descrito por una plantilla T y otro por una sub-plantilla S .

Teniendo en cuenta que no todos los parches en una sub-plantilla son igualmente importantes cuando su medida de consistencia local es sumada para obtener una medida de consistencia global, se deben filtrar aquellos parches que deberían tener un bajo peso en la medida global. Por ejemplo, si un parche espacio-temporal P_I no tiene movimiento alguno (la intensidad de todos sus

pixeles no cambia durante 3 fotogramas), al tomar una medida de consistencia local vemos que su matriz \mathbf{M}_1 posee derivadas temporales iguales a cero y por lo tanto su respectivo determinante también es igual a cero, lo que ocasiona una medida de consistencia demasiado grande en el caso práctico. Entonces, estos parches uniformes o sin cambio de movimiento (que generalmente hacen parte del fondo), aunque deberían tener un bajo peso, tienen una alta consistencia de movimiento con cualquier otro parche uniforme. Esto produce un alto puntaje de correlación en regiones uniformes donde no hay movimiento (por ejemplo, el fondo en un video), lo cual es indeseado para los resultados esperados en este proyecto. Por esta razón, el bloque de detección movimiento (explicado en la Sección 3.1.3.2), fue implementado en el sistema, permitiendo así una selección previa de parches espacio-temporales con medidas de consistencia local válidas.

Ahora sí, habiendo mencionado todo lo anterior, se prosigue con la definición de la medida de consistencia global entre una plantilla T y una sub-plantilla S . Para obtener esta medida, [5] sugiere computar el valor promedio de $1/m_{TS}$; esto es:

$$\frac{1}{N} \sum \frac{1}{m_{TS}} \quad (23)$$

donde N es el número de parches espacio-temporales de S .

4.1.3.4.1 Puntaje global para S (modificación a [5])

Debido a la existencia de parches uniformes en casi todas las plantillas y sub-plantillas (parches pertenecientes al fondo de estas secuencias de video), la medida de consistencia global planteada por los autores del presente trabajo de grado, consiste en computar el valor promedio de la medida de consistencia local $1/m_{TS}$, pero únicamente sobre el número de parches espacio-temporales no uniformes de la plantilla T . A saber, cuando la plantilla se compara con ella misma, todos los parches no uniformes en ésta producen una medida de consistencia local (que está entre 0.0001 y 1), esto es:

$$\frac{1}{N_T} \sum \frac{1}{m_{TS}} \quad (24)$$

Donde N_T es el número de parches espacio-temporales no uniformes de T .

Si la medida de consistencia global se hiciera computando el valor promedio de $1/m_{TS}$ sobre el número de parches no uniformes de la sub-plantilla S (esto es N_S), esto es:

$$\frac{1}{N_S} \sum \frac{1}{m_{TS}} \quad (25)$$

Esta medida obtenida podría ser bastante alta (cercana a 1) cuando la sub-plantilla S posea algunos pocos parches espacio-temporales no uniformes (o de movimiento) y consistentes con sus parches correspondientes en la plantilla T . En la Figura 14, se observan algunos resultados generados con esta medida (ecuación (23)), donde existen sub-plantillas con un alto valor de consistencia global, debido a unos pocos parches aislados con puntajes altos de consistencia local.



Figura 14. Algunos resultados obtenidos de computar la medida de consistencia global con el número de parches no uniformes de la sub-plantilla S (N_S). Se observan muchos parches dispersos que, aunque consistentes con parches de la plantilla T , producen un resultado demasiado confuso.

Por lo tanto, la medida de consistencia global definitiva en este proyecto, será la obtenida en la ecuación (24), resultado de computar el valor promedio de $1/m_{TS}$ sobre el número de parches espacio-temporales de la plantilla T que poseen un alto puntaje de consistencia local cuando ésta se compara con ella misma. En este caso, los parches con un puntaje de consistencia alto son todos aquellos que son no uniformes, es decir, aquellos que contienen algún movimiento.

4.1.4 Detección mayor puntaje de correlación (modificación a [5])

El objetivo de este trabajo de grado es desarrollar un algoritmo que permita encontrar en un video de entrada, el movimiento más parecido al descrito en otra secuencia de video (generalmente más corta). En otras palabras, se pretende encontrar en un video V , la sub-plantilla S con la medida de consistencia más alta al compararla con una plantilla T . Siendo así, una vez el sistema procesa una nueva sub-plantilla S , resultado de realizar el proceso descrito en la sección 3.1.1, procede a

comparar la nueva medida de consistencia global (o valor del puntaje de correlación de la sub-plantilla) con el valor del puntaje de correlación de la sub-plantilla anterior. Si la nueva medida es mayor que la antigua, el sistema descarta la vieja y almacena las coordenadas espaciales y temporales de esta nueva sub-plantilla, que posee (hasta el momento) la medida más alta de consistencia global con la plantilla T . Si por el contrario la nueva medida es menor, el sistema descarta la nueva sub-plantilla y procede a analizar otra.

De esta manera, el sistema continúa procesando sub-plantillas a medida que recorre todo el video V , y siempre que encuentre alguna con un valor de puntaje de correlación más alto que la de mayor puntaje de correlación hasta ese momento, el sistema descartará la antigua y seleccionará temporalmente la nueva sub-plantilla como la de mayor puntaje de correlación. Una vez el video a analizar fue recorrido en su totalidad, el sistema obtiene las coordenadas espaciales y temporales de la sub-plantilla con el puntaje de correlación más alto; esto es, en qué parte del video V se encuentra la sub-plantilla S que contiene el movimiento más parecido al descrito por la plantilla T .

Para almacenar (temporalmente) la ubicación de la sub-plantilla con mayor puntaje de correlación, el sistema procede a realizar un procesamiento similar al descrito en la sección 3.1.3, pero con dos diferencias específicas. Este procesamiento consta de varios bloques que se muestran en la Figura 15.

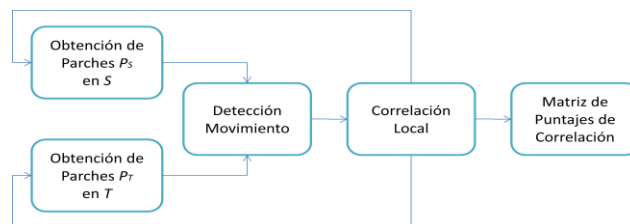


Figura 15. Diagrama en bloques que describe el proceso para encontrar las coordenadas de la sub-plantilla con mayor puntaje de correlación.

La primera diferencia está en que las secuencias que analiza, por un lado la plantilla T y por el otro la sub-plantilla S con mayor Puntaje de correlación (hasta ese momento), no están diezmadas, es decir que contienen más parches espacio-temporales. Por lo tanto, el análisis de los bloques de “Detección movimiento” y de “Correlación local” es más exhaustivo y por

consiguiente más costoso computacionalmente. Esto con el fin de realizar un análisis más detallado y sin pérdidas de información, para obtener un resultado más preciso.

La segunda diferencia está en la ausencia del bloque de “Medida de consistencia global” y en la existencia de un nuevo bloque llamado “Matriz de puntajes de correlación”. La ausencia del bloque descrito en la sección 3.1.3.4, es debida a que ya no es necesario computar el valor del puntaje de correlación de la sub-plantilla S , pues ya se tiene. Por otro lado, la existencia del nuevo bloque es debida a la necesidad de almacenar los datos relevantes de la sub-plantilla (con mayor puntaje de correlación hasta ese momento) para posteriormente generar el video resultado R . A saber, los datos relevantes son los siguientes:

- Coordenadas espaciales y temporales de cada parche espacio-temporal no uniforme de la sub-plantilla S , y su medida de consistencia local con su respectivo parche espacio-temporal de la plantilla T .
- Coordenadas espaciales y temporales de la sub-plantilla S (con mayor puntaje de correlación hasta ese momento), y su medida de consistencia global con la plantilla T (esto es, su valor de puntaje de correlación).

4.2 Creación del Video Resultado R (modificación a [5])

Por último, una vez obtenido la sub-plantilla S con la medida más alta de consistencia global con la plantilla T , se procede a generar el video resultado R .

El primer paso en este proceso, consiste en el ordenamiento de la matriz de puntajes de correlación que facilita el análisis de los datos existentes en ésta, y proporciona al algoritmo implementado una mayor eficiencia al crear el video resultado R .

El ordenamiento de la matriz se realiza de acuerdo a las coordenadas temporales de cada parche, permitiendo así combinar la información, del video original V y la matriz de puntajes de correlación, fotograma a fotograma.

Básicamente, este nuevo video R es el mismo video V en grises; pero, adicionalmente, en los fotogramas que contienen el movimiento más parecido al descrito por la plantilla T (esto es la sub-plantilla S con mayor puntaje de correlación), los parches espacio-temporales que

corresponden a los de la sub-plantilla S elegida están pintados de acuerdo a las siguientes convenciones:

- Si el valor de la medida de consistencia local de dicho parche espacio-temporal con su respectivo parche en la plantilla T es mayor a 0.8 (recordemos que esta medida está limitada entre 0 y 1), entonces este mismo parche espacio-temporal se pintará de color rojo en el video original que se encuentra en grises.
- Ahora, si el valor de la medida de consistencia local de dicho parche espacio-temporal es mayor a 0.6 y menor a 0.8, entonces este mismo parche espacio-temporal se pintará de color amarillo.
- Por otro lado, si el valor de la medida de consistencia local de dicho parche espacio-temporal es mayor a 0.4 y menor a 0.6, entonces este mismo parche espacio-temporal se pintará de color cian.
- Por último, si el valor de la medida de consistencia local de dicho parche espacio-temporal es menor a 0.4, entonces este mismo parche espacio-temporal no se modificará y por lo tanto se dejará en grises.

Después de generar fotograma a fotograma el video resultado R , el video es mostrado al usuario quien procede a analizar los resultados obtenidos. En la figura 16 se observan algunos fotogramas de un video resultado.

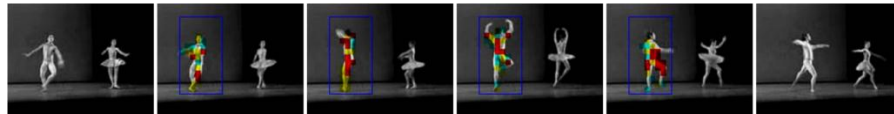


Figura 16. Algunos fotogramas de un video resultado R .

5. ANÁLISIS DE RESULTADOS

A continuación se describe el protocolo de pruebas utilizado para evaluar el desempeño del algoritmo desarrollado en este trabajo de grado.

5.1 Pruebas de Calibración

Las pruebas de calibración se realizaron con el video de una presentación de la compañía Birmingham Royal Ballet del Reino Unido. Para el análisis de éste se utilizó como plantilla un giro específico de uno de los bailarines en el video para analizar la robustez del sistema a la complejidad de los movimientos. En la Figura 17 se observan algunos fotogramas del video largo V y de la plantilla T (que fue sacada de V).

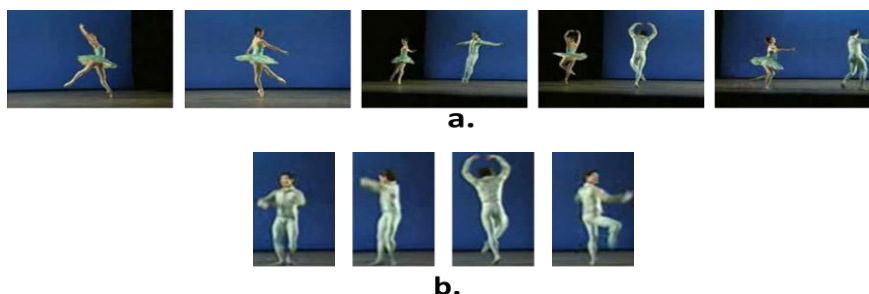


Figura 17. fotogramas de una presentación de la compañía Birmingham Royal Ballet del reino Unido. a. video largo V . b. plantilla T .

Debido a la larga duración del video de entrada V (a saber, 766 fotogramas) y para evitar un muy elevado tiempo de procesamiento por parte del sistema, las pruebas, una a una, fueron realizadas con segmentos más cortos de este mismo video, escogiendo fragmentos del video que contuvieran ciertos movimientos de interés, tanto similares como diferentes al descrito por la plantilla T .

Los resultados obtenidos durante las pruebas, están compuestos por un video resultado R que se especificó en la sección 3.3.2, y por una tabla de resultados numéricos donde se muestran algunos parámetros influyentes en la determinación de la sub-plantilla S que contiene el movimiento más parecido al descrito por la plantilla T . Estos parámetros son explicados a continuación:

- **Factor de diezmado:** Indica en qué nivel de resolución se está analizando el video, de acuerdo al factor de diezmado escogido por el usuario al inicio de la prueba. Para las

pruebas de calibración este parámetro se escogió como 3, indicando este número que el video original V se va a diezmar 2 veces en total. Específicamente, en el primer ciclo de diezmo (factor 3) la resolución del video a analizar es la cuarta parte de la del video original; asimismo, para el siguiente ciclo de diezmo (factor 2) la resolución del video a analizar es la mitad de la del video original; y finalmente, para el último ciclo de diezmo (factor 1) el video a analizar tiene la misma resolución del video original.

- **Puntaje global de T :** Indica el puntaje de consistencia global de la plantilla T al compararla con ella misma.
- **Parches N_T :** Indica el número de parches espacio-temporales no uniformes de T , es decir, el número de parches que hacen parte del movimiento descrito en esta secuencia de video. Para la plantilla, el número de parches no uniformes es el mismo número de parches que poseen un puntaje alto de consistencia local cuando ésta se compara con sí misma.
- **Puntaje global de S :** Indica el puntaje de consistencia global de una sub-plantilla S al compararla con la plantilla T . Para las pruebas realizadas, el puntaje global que se muestra en la tabla de resultados es el puntaje de la sub-plantilla S más parecida a T entre todas las sub-plantillas analizadas durante un ciclo de diezmo.
- **Parches N_S :** Indica el número de parches espacio-temporales no uniformes de S , es decir, el número de parches que hacen parte del movimiento descrito en esta secuencia de video. Cabe recordar que, como se mencionó en la sección 3.1.3.4, la medida de consistencia global entre la plantilla T y una sub-plantilla S consiste en computar el valor promedio de la medida de consistencia local, sobre el número de parches espacio-temporales no uniformes de la plantilla T (N_T), y no sobre los de S (N_S).
- **Parches color rojo:** Indica el número de parches espacio-temporales no uniformes de S , pero que a su vez tiene una medida de consistencia local (con su respectivo parche en T) mayor a 0.8. Estos son los parches que presentan mayor consistencia en su movimiento.
- **Parches color amarillo:** Indica el número de parches espacio-temporales no uniformes de S , pero que a su vez tiene una medida de consistencia local (con su respectivo parche en T) mayor a 0.6 y menor a 0.8.

- **Parches color cian:** Indica el número de parches espacio-temporales no uniformes de S , pero que a su vez tiene una medida de consistencia local (con su respectivo parche en T) mayor a 0.4 y menor a 0.6.
- **Otros parches:** Indica el número de parches espacio-temporales no uniformes de S , pero que a su vez tiene una medida de consistencia local (con su respectivo parche en T) menor a 0.4. Estos son los parches menos consistentes.
- **X inicial y Y inicial:** Indica las coordenadas espaciales del primer pixel de la sub-plantilla S con respecto a su ubicación en el video a analizar.
- **Fotograma inicial:** Indica la coordenada temporal del primer pixel de la sub-plantilla S con respecto a su ubicación en el video a analizar. Es decir, indica cuál es el primer fotograma de S con respecto a los fotogramas del video a analizar.

Todas las pruebas de calibración se realizaron con un factor de diezmado igual a 3 y un parámetro de vecindad igual a 3. Este último (que ingresa el usuario antes de la prueba) indica, como se mencionó en la sección 3.1.2, cuántos pixeles en todas las dimensiones y direcciones (espacio-tiempo) se van a analizar alrededor de las coordenadas X inicial, Y inicial y fotograma inicial de una sub-plantilla. A manera de ejemplo, si el parámetro vecindad es 3, implica que la región de análisis en un ciclo de diezmado (a excepción del primer ciclo) empezará 3 pixeles antes (en espacio y tiempo) que el pixel inicial de la sub-plantilla S con mayor puntaje global en el ciclo de diezmado anterior. Asimismo, la región de análisis terminará 3 pixeles después (en espacio y tiempo) que el pixel final de la sub-plantilla con mayor puntaje global en el ciclo de diezmado previo. La Figura 19 muestra de manera gráfica lo mencionado anteriormente.

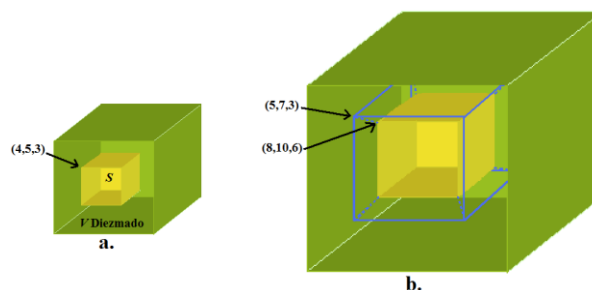


Figura 18. Región de Análisis dada por una Vecindad al cambiar de un ciclo de diezmado a otro. a. Primer ciclo de diezmado y sub-plantilla S con mayor puntaje Global. **b.** Siguiendo ciclo de diezmado y región de análisis alrededor de la sub-plantilla encontrada en a.

5.1.1 Comparación de la plantilla T con ella misma

Escogiendo como video a analizar V el mismo segmento de video correspondiente a la plantilla T , los resultados obtenidos después de procesar los dos videos se muestran en la Figura 20 y en la Tabla 1.

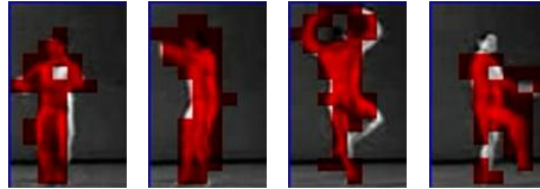


Figura 19. Algunos fotogramas del Video Resultado al comparar la plantilla T con ella misma.

FACTOR DE DIEZMADO: 3		FACTOR DE DIEZMADO: 2		FACTOR DE DIEZMADO: 1		
Puntaje Global de T	1.00	Puntaje Global de T	1.00	Puntaje Global de T	1.00	
parches N_T	10	parches N_T	81	parches N_T	293	
Puntaje Global de S	1.00	Puntaje Global de S	1.00	Puntaje Global de S	1.00	
parches N_S	10	parches N_S	81	parches N_S	293	
parches color Rojo	10	parches color Rojo	81	parches color Rojo	293	
parches color Amarillo	0	parches color Amarillo	0	parches color Amarillo	0	
parches color Cian	0	parches color Cian	0	parches color Cian	0	
Otros parches	0	Otros parches	0	Otros parches	0	
X Inicial	0	X Inicial	0	X Inicial	0	
Y Inicial	0	Y Inicial	0	Y Inicial	0	
fotograma Inicial	0	fotograma Inicial	0	fotograma Inicial	0	
					Tiempo de Proceso	24 seg

Tabla 1. Resultados numéricos al comparar la plantilla T con ella misma.

De la Tabla 1 se puede observar que el puntaje global obtenido en los 3 ciclos de diezmo es aproximadamente 1 (realmente es un poco mayor debido a ε en la ecuación 22) y que el número de parches espacio-temporales no uniformes de T (N_T) es igual al de S (N_S). Asimismo, se observa que en la sub-plantilla S únicamente existen parches de color rojo y por tanto existen tantos parches rojos como parches de movimiento N_S . Por esta razón, la Figura 19 sólo contiene fotogramas con parches espacio-temporales de color rojo.

5.1.2 Desplazamiento espacial por parches alrededor de la plantilla

Las primeras pruebas realizadas del algoritmo, se efectuaron con un desplazamiento espacial por parches espacio-temporales. A saber, el desplazamiento entre una sub-plantilla S y otra, era de 1

parche en tiempo y en espacio, lo cual arrojaba resultados con detecciones incorrectas del movimiento descrito por la plantilla. A manera de ejemplo, si un parche espacio-temporal está compuesto por 3 fotogramas cada uno de 7 pixeles de alto y 7 pixeles de ancho ($7 \times 7 \times 3$), la ubicación de una nueva sub-plantilla estaba determinada por estas dimensiones y por lo tanto una sub-plantilla adyacente a otra estaba ubicada 7 pixeles (más arriba, más abajo, más a la derecha o más a la izquierda) y/o 3 fotogramas (más atrás o más adelante) al lado.

Debido a los resultados incorrectos obtenidos al escoger sub-plantillas desplazadas espacialmente por parches espacio-temporales, se realizó un análisis espacial donde se desplazó la plantilla T en saltos de dimensión espacial igual al tamaño de un parche, alrededor de un movimiento exactamente igual a la plantilla, en una vecindad de 2 parches a lo alto y 4 parches a lo ancho (el análisis realizado no fue simétrico porque el movimiento estaba descrito más hacia lo alto que hacia lo ancho, y por tanto no cabían más sub-plantillas en el video V). La Gráfica 1 muestra la variación de los puntajes de consistencia global a medida que las sub-plantillas cambian parche a parche.

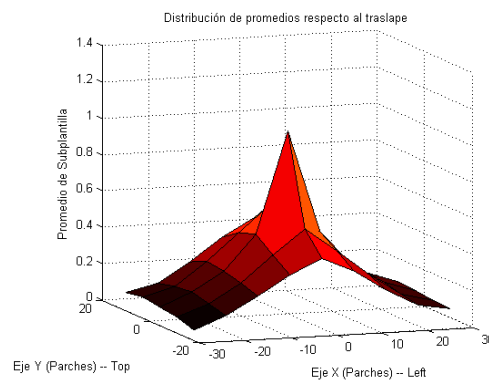


Figura 20. Distribución de los puntajes globales respecto al desplazamiento espacial parche a parche alrededor de una plantilla.

De la Figura 20 se observa que los puntajes de consistencia global caen rápidamente a medida que se presenta un desplazamiento relativo por parches entre las sub-plantillas S alrededor de la sub-plantilla con mayor puntaje (aquella que describe el movimiento exacto de la plantilla). A saber, el siguiente puntaje global más alto, ubicado en los parches inmediatamente contiguos al punto (0,0), tiene un valor aproximadamente igual a 0.44, presentando una caída de más del 50% con un desplazamiento de tan sólo 1 parche.

5.1.3 Desplazamiento espacial por pixeles alrededor de la plantilla

Al observar que la distancia relativa entre sub-plantillas no debe ser de 1 parche debido a la sensibilidad mostrada por la Figura 20, se procede a analizar la sensibilidad del resultado obtenido debido a desplazamientos relativos pixel a pixel alrededor de un movimiento exactamente igual al descrito por la plantilla T . Por lo tanto, se realizó un análisis espacial donde se desplazó la plantilla T en una vecindad de 12 pixeles alrededor de la misma plantilla pero contenida en el video V . En la Figura 21, se observa el comportamiento que produce la variación de los puntajes globales en función del desplazamiento espacial pixel a pixel de una plantilla.

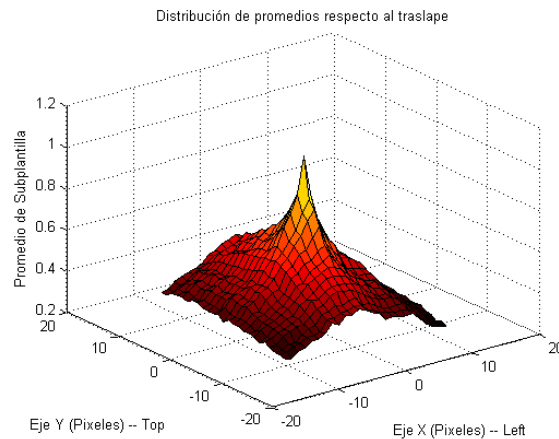


Figura 21. Distribución de los puntajes globales respecto al desplazamiento espacial pixel a pixel alrededor de una plantilla.

Como se observa en la Figura 21, el puntaje de consistencia global más alto, que está centrado en $(0,0)$, tiene un valor aproximadamente igual a 1, haciendo referencia a la coincidencia perfecta de la plantilla T con una sub-plantilla S exactamente idéntica. El siguiente puntaje global más alto, ubicado en los 8 pixeles inmediatamente contiguos al punto $(0,0)$, tiene un valor aproximadamente igual a 0.81, siendo este un puntaje relativamente alto aún. Para los 7 pixeles adyacentes consecutivos, el puntaje global aproximado es: 0.75, 0.68, 0.6, 0.54, 0.5, 0.44 y 0.4 respectivamente.

Por tanto, se concluye que el algoritmo es bastante sensible al desplazamiento relativo parche a parche, pero no tanto al desplazamiento espacial pixel a pixel; y por lo tanto se pueden lograr detecciones correctas del movimiento siempre y cuando la distancia relativa entre sub-plantillas no sea demasiado grande. Sin embargo, al ser la búsqueda más exhaustiva en un desplazamiento

por píxeles, el algoritmo se hace mucho más lento y en consecuencia el tiempo de procesamiento es mucho más alto. Por tal motivo existe un compromiso entre la precisión espacial con la que se quiere encontrar el movimiento y el costo computacional.

5.1.4 Desplazamiento temporal alrededor de la plantilla

Para analizar la sensibilidad del resultado obtenido debido a la localización temporal de la sub-plantilla alrededor de un movimiento exactamente igual al descrito por la plantilla T , se realizó un análisis temporal donde se desplazó la plantilla T en una vecindad de 12 fotogramas alrededor de la misma plantilla pero contenida en el video V . En la Figura 22, se observa el comportamiento obtenido con la variación de los puntajes globales en función del desplazamiento temporal de una plantilla..

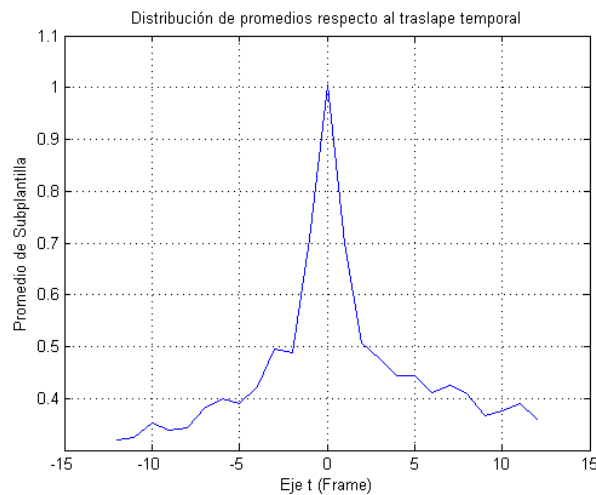


Figura 22. Distribución de los puntajes globales respecto al desplazamiento temporal fotograma a fotograma alrededor de una plantilla.

En la Figura 22 se observa que el algoritmo es más sensible a variaciones temporales que a variaciones espaciales (si se compara con los resultados mostrados en la Gráfica 2). Específicamente, el puntaje global de una sub-plantilla S que tiene 1 fotograma de desfase respecto a la plantilla T es aproximadamente de 0.7. Por otro lado, cuando el desplazamiento temporal relativo es de 2, el puntaje de consistencia global entre los dos segmentos de video es de aproximadamente 0.5.

En la misma línea del análisis anterior, se procedió a realizar una búsqueda temporal en un video relativamente largo (140 fotogramas) que contiene un movimiento idéntico al descrito por la

plantilla T , pero donde el fotograma inicial de la sub-plantilla que contiene dicho movimiento, se pierde al realizar el diezmado temporal (con factor de diezmado igual a 3). A saber, el fotograma inicial de la sub-plantilla S , que es idéntica a la plantilla T , corresponde al fotograma No. 71 en el video a analizar V . Por tanto, en el primer ciclo de diezmado (factor 3) la información contenida en el fotograma No. 71 se pierde pues 71 no es divisible por 4. En la Figura 23 se muestran algunos fotogramas del resultado obtenido. Se observa que aunque el video contiene el movimiento de la plantilla, la detección no es correcta. En la Tabla 2 se muestran los resultados numéricos obtenidos.



Figura 23. Algunos fotogramas del video resultado al buscar temporalmente la plantilla T en un video que contiene un movimiento idéntico pero en una posición temporal inicial no divisible por 4.

FACTOR DE DIEZMADO: 3		FACTOR DE DIEZMADO: 2		FACTOR DE DIEZMADO: 1	
Puntaje Global de T	1.00	Puntaje Global de T	1.00	Puntaje Global de T	1.00
parches N_T	4	parches N_T	46	parches N_T	286
Puntaje Global de S	0.93	Puntaje Global de S	0.54	Puntaje Global de S	0.33
parches color Rojo	4	parches color Rojo	16	parches color Rojo	35
parches color Amarillo	0	parches color Amarillo	11	parches color Amarillo	59
parches color Cian	0	parches color Cian	6	parches color Cian	27
Otros parches	0	Otros parches	3	Otros parches	48
X Inicial	0	X Inicial	0	X Inicial	0
Y Inicial	0	Y Inicial	0	Y Inicial	0
fotograma Inicial	0	fotograma Inicial	1	fotograma Inicial	2
				Tiempo de Proceso	45 seg

Tabla 2. Resultados numéricos de una detección incorrecta al buscar temporalmente la plantilla T en un video que contiene un movimiento idéntico pero en una posición temporal inicial no divisible por 4.

Cómo se observa claramente en la Tabla 2, la sub-plantilla con puntaje de consistencia global más alto con respecto a la plantilla (diezmada en un factor 3), tiene su fotograma inicial en 0, lo que equivale también a 0 en el video original (a saber, $0*4=0$). Sin embargo, cómo se conoce previamente que el fotograma inicial de la sub-plantilla que es idéntica a la plantilla está ubicado en el fotograma No. 71 del video original, se asume que en el primer ciclo de diezmado (factor 3)

la detección correcta debería estar entre los fotogramas 17 y 18 (a saber, $71/4=17.75$). En la Tabla 3 se observa el resultado numérico obtenido para los fotogramas 17 y 18.

FACTOR DE DIEZMADO: 3	
Puntaje Global de T	1.00
parches Nr	4
Puntaje Global de S	0.87
parches color Rojo	3
parches color Amarillo	1
parches color Cian	0
Otros parches	0
X Inicial	0
Y Inicial	0
fotograma Inicial	17

a.

b.

Tabla 3. Resultados numéricos para las sub-plantillas más cercanas a la sub-plantilla correcta (factor de diezmado igual a 3) a. sub-plantilla con fotograma inicial 17. **b.** sub-plantilla con fotograma inicial 18.

De los resultados mostrados en la Tabla 3 se entiende el por qué la detección fue incorrecta. Allí, se observa cómo el puntaje global para ambas sub-plantillas es menor al puntaje encontrado en la Tabla 2 (factor de diezmado 3). Esto ocurre debido a la pérdida de información al diezmarse temporalmente el video original, lo cual se espera de acuerdo a los resultados encontrados en la Figura 23, donde se observa que la sensibilidad del algoritmo al desplazamiento entre fotogramas es extremadamente alta, y por tanto, cuando se realiza un diezmado del video original (por ejemplo, un factor de diezmado igual a 3) la información temporal que se pierde en cada ciclo de diezmado es mucho más influyente en la obtención de una detección incorrecta que la información espacial perdida.

Por esta razón, se decidió que el diezmado temporal es siempre menor que el diezmado espacial en el algoritmo; específicamente, si el factor de diezmado es 3, en el primer ciclo de diezmado (factor 3) la resolución temporal del video a analizar es la mitad de la del video original; y para los dos siguientes ciclos de diezmado (factor 2 y factor 1, respectivamente) la resolución temporal del video a analizar es la misma del video original.

Al realizar esta implementación y corrección en el algoritmo, se obtienen los resultados mostrados en la Figura 24 y la Tabla 4.



Figura 24. Algunos fotogramas del video resultado al buscar temporalmente la plantilla T en un video que contiene un movimiento idéntico. (El factor de diezmado temporal es menor que el espacial).

FACTOR DE DIEZMADO: 3		FACTOR DE DIEZMADO: 2		FACTOR DE DIEZMADO: 1	
Puntaje Global de T	1.00	Puntaje Global de T	1.00	Puntaje Global de T	1.00
parches N_T	10	parches N_T	85	parches N_T	286
Puntaje Global de S	1.00	Puntaje Global de S	1.00	Puntaje Global de S	1.00
parches color Rojo	10	parches color Rojo	85	parches color Rojo	286
parches color Amarillo	0	parches color Amarillo	0	parches color Amarillo	0
parches color Cian	0	parches color Cian	0	parches color Cian	0
Otros parches	0	Otros parches	0	Otros parches	0
X Inicial	0	X Inicial	0	X Inicial	0
Y Inicial	0	Y Inicial	0	Y Inicial	0
fotograma Inicial	35	fotograma Inicial	71	fotograma Inicial	71
				Tiempo de Proceso	94 seg

Tabla 4. Resultados numéricos de una detección correcta al buscar temporalmente la plantilla T en un video que contiene un movimiento idéntico. (El factor de diezmado temporal es menor que el espacial).

5.1.5 Búsqueda de un movimiento IDENTICO al de la plantilla

Al realizar una búsqueda espacio-temporal de un movimiento descrito por una plantilla T en un video que efectivamente contiene una sub-plantilla con un movimiento idéntico, se encontró que el algoritmo es bastante preciso y produce resultados bastante exactos en la detección de dicho movimiento. La Figura 25 muestra algunos fotogramas del video resultado.

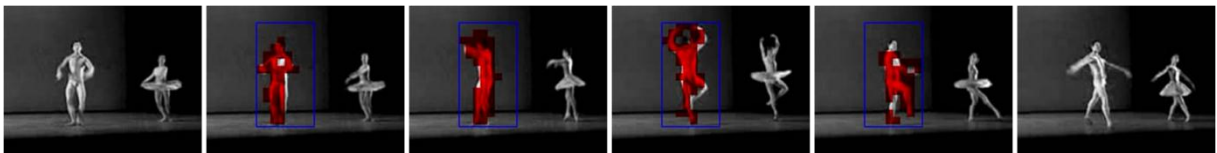


Figura 25. Algunos fotogramas del video resultado de la detección de un movimiento idéntico al descrito en T .

Allí se observa cómo el algoritmo después de analizar todo el video (en espacio y tiempo), realiza una detección de la sub-plantilla S con mayor puntaje global, que coincide exactamente con la plantilla T . En la Tabla 5 se muestran los resultados numéricos de las sub-plantillas con mayores

puntajes de consistencia global con respecto a cada ciclo de diezmado.

FACTOR DE DIEZMADO: 3		FACTOR DE DIEZMADO: 2		FACTOR DE DIEZMADO: 1	
Puntaje Global de T	1.00	Puntaje Global de T	1.01	Puntaje Global de T	1.00
parches N_T	8	parches N_T	75	parches N_T	278
Puntaje Global de S	0.92	Puntaje Global de S	0.78	Puntaje Global de S	1.00
parches color Rojo	7	parches color Rojo	58	parches color Rojo	278
parches color Amarillo	1	parches color Amarillo	2	parches color Amarillo	0
parches color Cian	0	parches color Cian	2	parches color Cian	0
Otros parches	0	Otros parches	3	Otros parches	0
X Inicial	4	X Inicial	7	X Inicial	13
Y Inicial	13	Y Inicial	24	Y Inicial	48
fotograma Inicial	36	fotograma Inicial	72	fotograma Inicial	72
				Tiempo de Proceso	5925 seg

Tabla 5. Resultados numéricos de la detección de un movimiento idéntico al descrito en T .

De los resultados contenidos en la Tabla 5, se observa que aunque en los dos primeros ciclos de diezmado (factor 3 y factor 2, respectivamente) el movimiento detectado en la sub-plantilla con mayor puntaje global, no corresponde exactamente al mismo movimiento descrito por la plantilla, el algoritmo realiza una búsqueda alrededor de las detecciones respectivas de cada ciclo y, finalmente, en el último ciclo de diezmado (es decir, en el video original) encuentra la sub-plantilla que contiene un movimiento idéntico al contenido en T . De esto se concluye que, tal como se mencionó en la sección 4.1.3, el algoritmo no es demasiado sensible a pérdidas de la información espacial entre ciclos de diezmado.

5.1.6 Búsqueda de un movimiento SIMILAR al de la plantilla

Como lo que se quiere no es detectar únicamente movimientos idénticos a los descritos por la plantilla sino también movimientos similares, es decir, sub-plantillas S que describan el mismo movimiento que T pero que no sean exactamente iguales a ésta, se proceden a realizar pruebas del algoritmo con videos que contienen movimientos muy parecidos pero no iguales al contenido en la plantilla espacio-temporal. La Figura 26 muestra algunos fotogramas del video resultado.

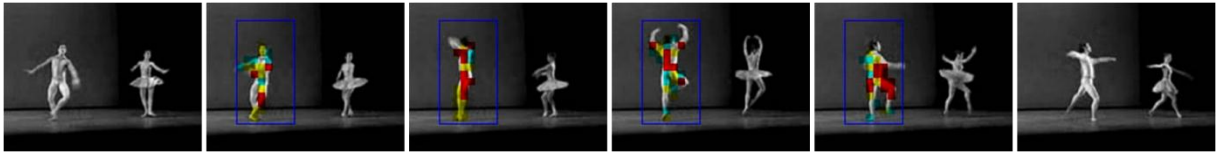


Figura 26. Algunos fotogramas del video resultado de la detección de un movimiento similar al descrito en T .

El sistema, después de analizar todo el video (en espacio y tiempo), realiza una detección de la sub-plantilla S con mayor puntaje global que, aunque no es exactamente la misma plantilla T , es la más parecida entre todas las posibles sub-plantillas en el video; y que efectivamente contiene un movimiento similar al descrito en T . En la Tabla 6 se muestran los resultados numéricos de las sub-plantillas con mayores puntajes de consistencia global con respecto a cada ciclo de diezmado.

FACTOR DE DIEZMADO: 3		FACTOR DE DIEZMADO: 2		FACTOR DE DIEZMADO: 1	
Puntaje Global de T	1.00	Puntaje Global de T	1.00	Puntaje Global de T	1.00
parches N_T	10	parches N_T	81	parches N_T	293
Puntaje Global de S	0.79	Puntaje Global de S	0.56	Puntaje Global de S	0.51
parches color Rojo	6	parches color Rojo	22	parches color Rojo	90
parches color Amarillo	4	parches color Amarillo	18	parches color Amarillo	86
parches color Cian	0	parches color Cian	11	parches color Cian	55
Otros parches	0	Otros parches	13	Otros parches	54
X Inicial	5	X Inicial	8	X Inicial	16
Y Inicial	8	Y Inicial	14	Y Inicial	28
fotograma Inicial	31	fotograma Inicial	61	fotograma Inicial	61
				Tiempo de Proceso	5643 seg

Tabla 6. Resultados numéricos de la detección de un movimiento similar al descrito en T .

De los resultados contenidos en la Tabla 6, se observa que aunque los puntajes globales de cada sub-plantilla no son muy altos (teniendo en cuenta que el mayor puntaje posible es aproximadamente 1; aun así el algoritmo logra detectar un movimiento similar al contenido en T , pues a pesar de la información perdida entre ciclos de diezmado, las sub-plantillas escogidas para cada ciclo tienen un puntaje de consistencia global mayor al de las demás sub-plantillas S que no contienen movimientos similares al descrito por la plantilla T .

5.1.7 Sensibilidad a la escala

Como lo que se quiere detectar es un movimiento específico independientemente de qué o quién lo realice, puede ocurrir que el movimiento de interés sea realizado por un objeto de menor o mayor tamaño al objeto que realiza ese mismo movimiento en la plantilla T . Para probar el desempeño del algoritmo implementado a cambios en la escala del movimiento, se realizaron algunas pruebas donde el movimiento contenido en la plantilla se encuentra contenido también en el video a analizar, pero en una escala menor. La Figura 27 muestra el video resultado de comparar la plantilla T con varias sub-plantillas S alrededor de un movimiento similar al descrito por T , pero realizado por un individuo de menor tamaño.

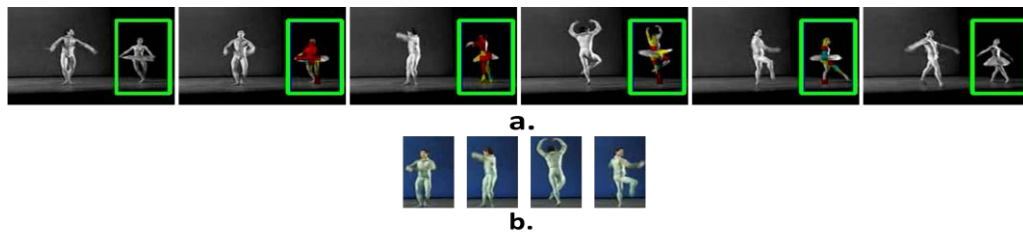


Figura 27. Video resultado de una búsqueda del movimiento contenido en T alrededor de un movimiento similar pero con una escala espacial menor. a. Algunos fotogramas del video resultado R . **b.** Algunos fotogramas de la plantilla T .

Como el video a analizar contiene dos movimientos similares al descrito por la plantilla T (a saber: uno idéntico, realizado por un hombre; y otro simplemente parecido, realizado por una mujer que es de menor tamaño que el hombre), si se analiza todo el video V el algoritmo detecta sólo el movimiento con mayor puntaje de consistencia global, como se muestra en la Figura 25. Sin embargo, como únicamente se quiere analizar el movimiento realizado por la mujer, se genera una región de interés alrededor de la mujer (enmarcada en la Figura 27a por el rectángulo verde) y se procede a comparar la plantilla con todas las sub-plantillas posibles en dicha región.

La Tabla 7 muestra los resultados numéricos para la sub-plantilla con más alto puntaje global dentro de la región de interés. El procesamiento se hizo únicamente para el último ciclo de diezmado (factor 1).

FACTOR DE DIEZMADO: 1	
Puntaje Global de T	1.00
parches N_T	293
Puntaje Global de S	0.34
parches N_S	202
parches color Rojo	47
parches color Amarillo	43
parches color Cian	39
Otros parches	59
X Inicial	20
Y Inicial	130
fotograma Inicial	69
Tiempo de Proceso	285 seg

Tabla 7. Resultados numéricos para una búsqueda del movimiento contenido en T alrededor de un movimiento similar pero con una escala espacial menor.

De los datos mostrados en la Tabla 7, se observa que el mayor puntaje global, correspondiente a la sub-plantilla S que contiene el movimiento más parecido al descrito en la plantilla T , es muy bajo comparado con el puntaje de la sub-plantilla escogida en el último ciclo de diezmado (factor 1) para la prueba realizada en la sección 4.1.5, mostrado en la Tabla 5. Además, este puntaje también es más bajo que el puntaje de la sub-plantilla escogida en el último ciclo de diezmado (factor 1) para la prueba realizada en la sección 5.1.6, mostrado en la Tabla 6. Por tanto, aunque el movimiento es similar en los tres casos, los resultados obtenidos muestran que el puntaje para la sub-plantilla detectada cuando la mujer realiza el movimiento, es significativamente menor. Si comparamos el número de parches no uniformes pertenecientes a la plantilla T (N_T), con el número de parches no uniformes pertenecientes a la sub-plantilla con mayor puntaje de consistencia global (N_S), se observa que N_S es mucho menor a N_T lo cual implica que el movimiento contenido en la secuencia de video S es mucho más pequeño que el movimiento contenido en T . De esto, se concluye que el algoritmo desarrollado en este trabajo de grado no es efectivo en la detección de un movimiento con una escala diferente al descrito por la plantilla. Se tiene entonces que, siendo el método utilizado variante a la escala del movimiento, para lograr una detección correcta es absolutamente necesario que el número de parches no uniformes en la

sub-plantilla S que contiene el movimiento más parecido al descrito por T , sea muy similar al número de parches no uniformes en esta última.

5.1.8 Búsqueda en videos sin un movimiento similar al de la plantilla

La última prueba de calibración se basa en observar los resultados numéricos de las detecciones respectivas en diferentes videos que no contienen un movimiento similar al de la plantilla T . Esto con el fin de identificar las diferencias entre los distintos parámetros de: los resultados generados a partir del análisis de una secuencia de video sin un movimiento similar a T , y los resultados generados a partir del análisis de una secuencia de video con un movimiento similar al descrito por T (resultados obtenidos en las secciones anteriores).

La Figura 28 muestra algunos fotogramas de una secuencia de video que no contiene ningún movimiento similar al de la plantilla T . También se muestran algunos fotogramas de la plantilla y del video resultado.

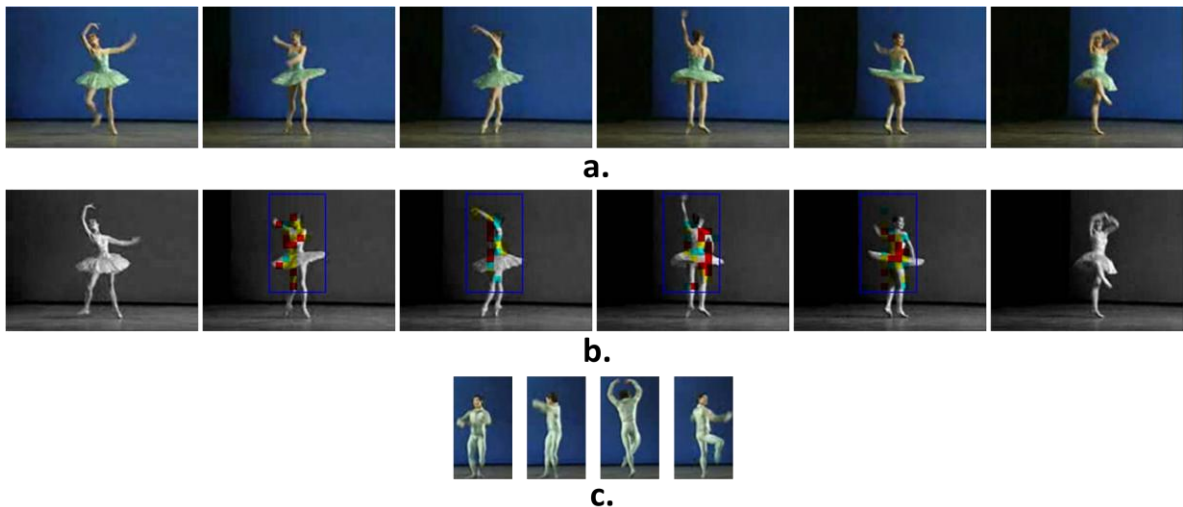


Figura 28. Análisis de una secuencia de video que no contiene ningún movimiento similar al contenido en T . **a.** Algunos fotogramas del video a analizar V . **b.** Algunos fotogramas del video resultado. **c.** Algunos fotogramas de la plantilla T .

En la Tabla 8 se muestran los resultados numéricos obtenidos en esta prueba.

FACTOR DE DIEZMADO: 3		FACTOR DE DIEZMADO: 2		FACTOR DE DIEZMADO: 1	
Puntaje Global de T	1.00	Puntaje Global de T	1.00	Puntaje Global de T	1.00
parches N_T	10	parches N_T	81	parches N_T	293

Puntaje Global de S	0.76	Puntaje Global de S	0.47	Puntaje Global de S	0.43
parches N_s	15	parches N_s	100	parches N_s	318
parches color Rojo	6	parches color Rojo	22	parches color Rojo	57
parches color Amarillo	2	parches color Amarillo	14	parches color Amarillo	62
parches color Cian	2	parches color Cian	13	parches color Cian	48
Otros parches	0	Otros parches	9	Otros parches	30
X Inicial	0	X Inicial	3	X Inicial	4
Y Inicial	16	Y Inicial	32	Y Inicial	66
fotograma Inicial	20	fotograma Inicial	38	fotograma Inicial	39
				Tiempo de Proceso	4143 seg

Tabla 8. Resultados numéricos de una búsqueda en un video que no contienen ningún movimiento similar al contenido en T .

De los resultados mostrados en la Tabla 8, se pueden hacer las siguientes comparaciones respecto a resultados anteriores:

- Para un movimiento que no es similar al movimiento descrito por la plantilla T , el puntaje de consistencia global de la sub-plantilla S que contiene dicho movimiento al compararlo con T , siempre es menor (respectivamente a cada ciclo de diezmo) al puntaje global de una sub-plantilla S que sí contiene un movimiento similar al de T (Vease la Tabla 6 y la Tabla 8).
- Para un movimiento que no es similar al descrito en la plantilla, el porcentaje de parches rojos y amarillos (que poseen los más altos puntajes de consistencia local) que resulta de comparar la sub-plantilla con la plantilla, generalmente es menor que el porcentaje de los mismos en un resultado proveniente de comparar una sub-plantilla con una plantilla, ambas con un movimiento similar. A saber, este porcentaje es respecto al número de parches espacio-temporales no uniformes pertenecientes a la plantilla T (N_T).
- Como se explicó en la sección 4.1.7, otro parámetro definitivo en la distinción de un movimiento similar al contenido en la plantilla de uno que no lo es, es el número de parches espacio-temporales no uniformes en la sub-plantilla S (N_S), esto debido a que el método implementado para la detección no es invariante a la escala. Por tanto, un movimiento con un tamaño mayor o menor al descrito por la plantilla T , es considerado distinto. En la Figura 28, se observa claramente que los individuos que realizan los movimientos en el video a analizar V y la plantilla T , son de tamaños distintos (esto es

debido a la perspectiva del video y a la posición relativa del sujeto con respecto a la cámara de video).

En la Tabla 9, se muestra la comparación específica para los resultados de las Tablas 6 y 8.

FACTOR DE DIEZMADO: 3		FACTOR DE DIEZMADO: 2		FACTOR DE DIEZMADO: 1	
parches N_T	10	parches N_T	81	parches N_T	293
Puntaje Global de S	0.76	Puntaje Global de S	0.47	Puntaje Global de S	0.43
parches N_S	15	parches N_S	100	parches N_S	318
parches Rojos (%)	60	parches Rojos (%)	27.16	parches Rojos (%)	19.45
parches Amarillos (%)	20	parches Amarillos (%)	17.28	parches Amarillos (%)	21.16

a.

FACTOR DE DIEZMADO: 3		FACTOR DE DIEZMADO: 2		FACTOR DE DIEZMADO: 1	
parches N_T	10	parches N_T	81	parches N_T	293
Puntaje Global de S	0.79	Puntaje Global de S	0.56	Puntaje Global de S	0.51
parches N_S	17	parches N_S	77	parches N_S	289
parches Rojos (%)	60	parches Rojos (%)	27.16	parches Rojos (%)	30.71
parches Amarillos (%)	40	parches Amarillos (%)	22.22	parches Amarillos (%)	29.35

b.

Tabla 9. Comparación de resultados entre los resultados de las Tablas 6 y 8. a. Resultados Tabla 8. b. Resultados Tabla 6

La diferencia más notable se observa en el último nivel de diezmado, donde ambas secuencias de video tienen mayor resolución.

5.2 Pruebas Generales

Después de calibrar el sistema y valorar su desempeño con un movimiento relativamente complejo como es un giro de ballet, se realizaron algunas pruebas generales con un movimiento un poco más sencillo, que pretendían evaluar la eficacia del algoritmo para detectar un movimiento específico.

Los videos utilizados en las pruebas generales, son videos que contiene tráfico vehicular y donde el movimiento de interés es un giro en L que realizan los vehículos al cruzar desde la Carrera 13 hacia la Calle 39. A saber, la plantilla T utilizada en las pruebas generales, contiene un giro en L realizado por un vehículo y se procede a comparar esta plantilla con todos los movimientos realizados por varios vehículos que se encuentran circulando a lo largo de un video V . En la

Figura 29 se observan algunos fotogramas de la plantilla mencionada. En la Tabla 10 se muestran las características de la plantilla *T*.



Figura 29. plantilla *T* de un vehículo realizando un giro en *L*

Ancho fotograma	56 pixeles
Alto fotograma	44 pixeles
Duración	00:00:01
fotogramas	20

Tabla 10. Características espacio-tiempo de la plantilla *T*.

Al comparar la plantilla con ella misma, se obtienen los resultados mostrados en la Tabla 11.

FACTOR DE DIEZMADO: 3		FACTOR DE DIEZMADO: 2		FACTOR DE DIEZMADO: 1		
Puntaje Global de <i>T</i>	1.00	Puntaje Global de <i>T</i>	1.00	Puntaje Global de <i>T</i>	1.00	
parches <i>N_r</i>	6	parches <i>N_r</i>	66	parches <i>N_r</i>	234	
Puntaje Global de <i>S</i>	1.00	Puntaje Global de <i>S</i>	1.00	Puntaje Global de <i>S</i>	1.00	
parches <i>N_s</i>	6	parches <i>N_s</i>	66	parches <i>N_s</i>	234	
parches color Rojo	6	parches color Rojo	66	parches color Rojo	234	
parches color Amarillo	0	parches color Amarillo	0	parches color Amarillo	0	
parches color Cian	0	parches color Cian	0	parches color Cian	0	
Otros parches	0	Otros parches	0	Otros parches	0	
X Inicial	0	X Inicial	0	X Inicial	0	
Y Inicial	0	Y Inicial	0	Y Inicial	0	
fotograma Inicial	0	fotograma Inicial	0	fotograma Inicial	0	
					Tiempo de Proceso	14 seg

Tabla 11. Resultados numéricos de comparar la plantilla *T* con sí misma.

Estos resultados son los esperados de acuerdo a las pruebas de calibración. El video resultado es demasiado pequeño espacialmente, y por esta razón Microsoft Visual C++ no puede generarlo.

A continuación se muestran varios resultados de algunas pruebas realizadas con diferentes videos, que están compuestos por:

- Algunos fotogramas del video a analizar *V*.
- Algunos fotogramas del video resultado *R*.

- Tabla con resultados numéricos de las sub-plantillas con mayor puntaje de consistencia global, para cada ciclo de diezmo.

En la Tabla 12 se muestran las características espacio-tiempo de los Videos utilizados en las siguientes pruebas.

PRUEBA	Ancho fotograma	Alto fotograma	Duración	fotogramas
Prueba No. 1	180 pixeles	144 pixeles	00:00:05	60
Prueba No. 2	180 pixeles	144 pixeles	00:00:04	52
Prueba No. 3	180 pixeles	144 pixeles	00:00:03	48
Prueba No. 4	180 pixeles	144 pixeles	00:00:04	52
Prueba No. 5	180 pixeles	144 pixeles	00:00:05	60

Tabla 12. Características espacio-tiempo de los Videos usados en las pruebas generales.

5.2.1 Prueba No. 1

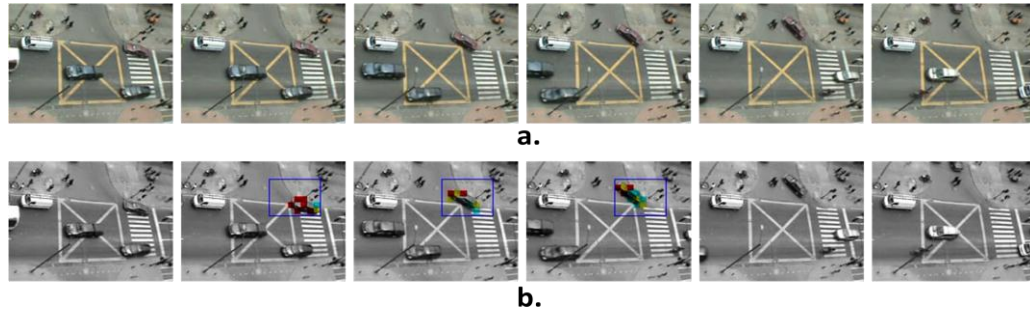


Figura 30. Videos de la Prueba No.1. a. Algunos fotogramas del video a analizar V. b. Algunos fotogramas del video resultado R.

FACTOR DE DIEZMADO: 3		FACTOR DE DIEZMADO: 2		FACTOR DE DIEZMADO: 1		
Puntaje Global de T	1.00	Puntaje Global de T	1.00	Puntaje Global de T	1.00	
parches $N\tau$	6	parches $N\tau$	66	parches $N\tau$	234	
Puntaje Global de S	0.88	Puntaje Global de S	0.64	Puntaje Global de S	0.56	
parches color Rojo	5	parches color Rojo	26	parches color Rojo	55	
parches color Amarillo	1	parches color Amarillo	17	parches color Amarillo	55	
parches color Cian	0	parches color Cian	8	parches color Cian	56	
Otros parches	0	Otros parches	14	Otros parches	50	
X Inicial	6	X Inicial	10	X Inicial	21	
Y Inicial	25	Y Inicial	48	Y Inicial	96	
fotograma Inicial	0	fotograma Inicial	2	fotograma Inicial	4	
					Tiempo de Proceso	1319 seg

Tabla 13. Resultados numéricos de la Prueba No.1

5.2.2 Prueba No. 2

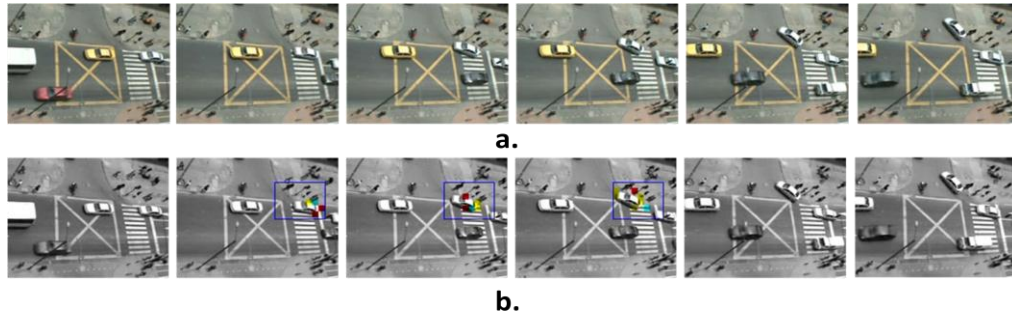


Figura 31. Videos de la Prueba No.2. a. Algunos fotogramas del video a analizar V. b. Algunos fotogramas del video resultado R.

FACTOR DE DIEZMADO: 3		FACTOR DE DIEZMADO: 2		FACTOR DE DIEZMADO: 1		
Puntaje Global de T	1.00	Puntaje Global de T	1.00	Puntaje Global de T	1.00	
parches N_T	6	parches N_T	66	parches N_T	234	
Puntaje Global de S	0.93	Puntaje Global de S	0.50	Puntaje Global de S	0.42	
parches color Rojo	6	parches color Rojo	18	parches color Rojo	32	
parches color Amarillo	0	parches color Amarillo	13	parches color Amarillo	51	
parches color Cian	0	parches color Cian	7	parches color Cian	45	
Otros parches	0	Otros parches	22	Otros parches	66	
X Inicial	6	X Inicial	13	X Inicial	29	
Y Inicial	29	Y Inicial	55	Y Inicial	109	
fotograma Inicial	12	fotograma Inicial	21	fotograma Inicial	18	
					Tiempo de Proceso	1316 seg

Tabla 14. Resultados numéricos de la Prueba No.2

5.2.3 Prueba No. 3

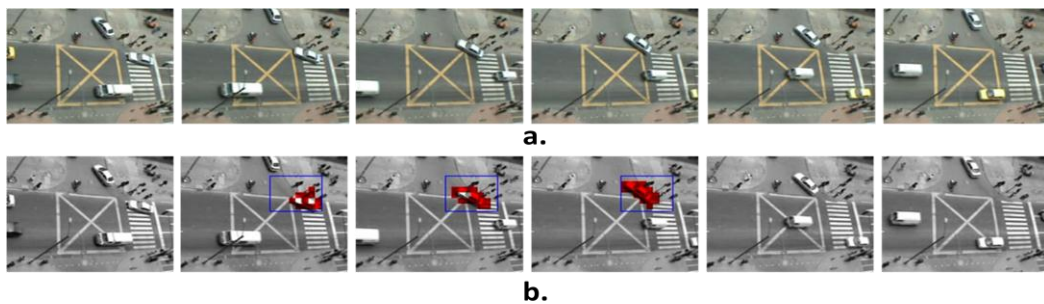


Figura 32. Videos de la Prueba No.3. a. Algunos fotogramas del video a analizar V. b. Algunos fotogramas del video resultado R.

FACTOR DE DIEZMADO: 3		FACTOR DE DIEZMADO: 2		FACTOR DE DIEZMADO: 1	
Puntaje Global de T	1.00	Puntaje Global de T	1.00	Puntaje Global de T	1.00
parches N_T	6	parches N_T	66	parches N_T	234
Puntaje Global de S	1.00	Puntaje Global de S	1.00	Puntaje Global de S	1.00
parches color Rojo	6	parches color Rojo	66	parches color Rojo	234
parches color Amarillo	0	parches color Amarillo	0	parches color Amarillo	0
parches color Cian	0	parches color Cian	0	parches color Cian	0
Otros parches	0	Otros parches	0	Otros parches	0
X Inicial	6	X Inicial	12	X Inicial	24
Y Inicial	24	Y Inicial	48	Y Inicial	96
fotograma Inicial	6	fotograma Inicial	12	fotograma Inicial	12
				Tiempo de Proceso	1257 seg

Tabla 15. Resultados numéricos de la Prueba No.3

5.2.4 Prueba No. 4

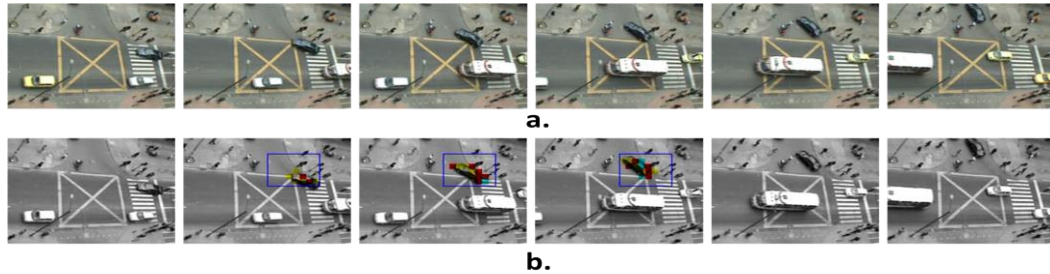


Figura 33. Videos de la Prueba No.4. a. Algunos fotogramas del video a analizar V. b. Algunos fotogramas del video resultado R.

FACTOR DE DIEZMADO: 3		FACTOR DE DIEZMADO: 2		FACTOR DE DIEZMADO: 1	
Puntaje Global de T	1.00	Puntaje Global de T	1.00	Puntaje Global de T	1.00
parches N_T	6	parches N_T	66	parches N_T	234
Puntaje Global de S	0.92	Puntaje Global de S	0.58	Puntaje Global de S	0.51
parches color Rojo	6	parches color Rojo	24	parches color Rojo	52
parches color Amarillo	0	parches color Amarillo	11	parches color Amarillo	60
parches color Cian	0	parches color Cian	13	parches color Cian	40
Otros parches	0	Otros parches	13	Otros parches	48
X Inicial	5	X Inicial	9	X Inicial	21
Y Inicial	21	Y Inicial	14	Y Inicial	90
fotograma Inicial	10	fotograma Inicial	19	fotograma Inicial	17
				Tiempo de Proceso	1120 seg

Tabla 16. Resultados numéricos de la Prueba No.4

5.2.5 Prueba No. 5

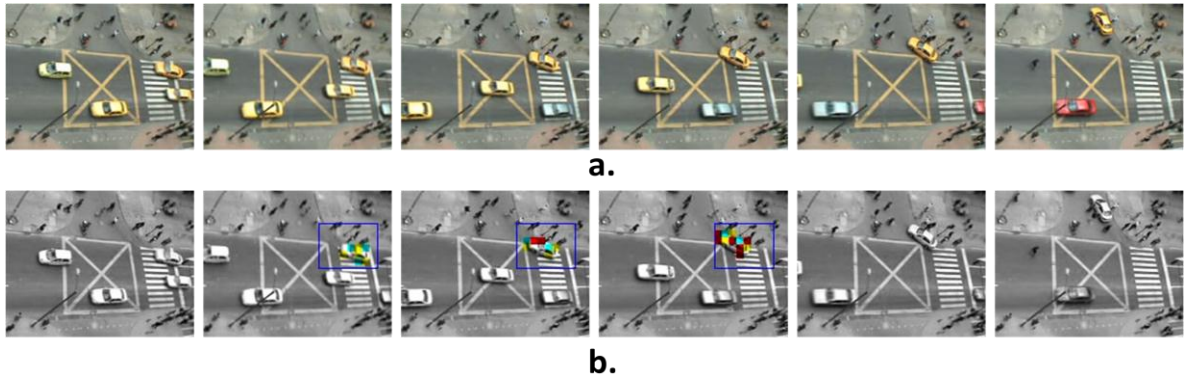


Figura 34. Videos de la Prueba No.5. a. Algunos fotogramas del video a analizar V. b. Algunos fotogramas del video resultado R.

FACTOR DE DIEZMADO: 3		FACTOR DE DIEZMADO: 2		FACTOR DE DIEZMADO: 1	
Puntaje Global de T	1.00	Puntaje Global de T	1.00	Puntaje Global de T	1.00
parches Nr	6	parches Nr	66	parches Nr	234
Puntaje Global de S	0.92	Puntaje Global de S	0.63	Puntaje Global de S	0.54
parches color Rojo	6	parches color Rojo	24	parches color Rojo	38
parches color Amarillo	0	parches color Amarillo	17	parches color Amarillo	71
parches color Cian	0	parches color Cian	13	parches color Cian	56
Otros parches	0	Otros parches	9	Otros parches	61
X Inicial	9	X Inicial	16	X Inicial	32
Y Inicial	29	Y Inicial	55	Y Inicial	110
fotograma Inicial	5	fotograma Inicial	9	fotograma Inicial	9
				Tiempo de Proceso	1992 seg

Tabla 17. Resultados numéricos de la Prueba No.5

5.2.6 Análisis de pruebas generales

De acuerdo a los resultados obtenidos de las pruebas realizadas en las Secciones 4.2.1 a 4.2.5, se puede concluir que el sistema funciona bastante bien, puesto que logra su objetivo con un porcentaje bastante alto de acierto.

En la Tabla 18, se muestra la comparación específica para los resultados de las Tablas 13 a 17, en donde se muestran los resultados numéricos de cada prueba.

FACTOR DE DIEZMADO: 1 Prueba No. 1		FACTOR DE DIEZMADO: 1 Prueba No. 2		FACTOR DE DIEZMADO: 1 Prueba No. 3		FACTOR DE DIEZMADO: 1 Prueba No. 4		FACTOR DE DIEZMADO: 1 Prueba No. 5	
Puntaje Global de <i>T</i>	1.00	Puntaje Global de <i>T</i>	1.00	Puntaje Global de <i>T</i>	1.00	Puntaje Global de <i>T</i>	1.00	Puntaje Global de <i>T</i>	1.00
parches <i>N_T</i>	234	parches <i>N_T</i>	234	parches <i>N_T</i>	234	parches <i>N_T</i>	234	parches <i>N_T</i>	234
Puntaje Global de <i>S</i>	0.55	Puntaje Global de <i>S</i>	0.42	Puntaje Global de <i>S</i>	1.00	Puntaje Global de <i>S</i>	0.51	Puntaje Global de <i>S</i>	0.54
Parches Rojos (%)	23,50	Parches Rojos (%)	13,67	Parches Rojos (%)	100	Parches Rojos (%)	22,22	Parches Rojos (%)	16,24
Parches Amarillos (%)	23,50	Parches Amarillos (%)	21,79	Parches Amarillos (%)	0	Parches Amarillos (%)	25,64	Parches Amarillos (%)	30,34

Tabla 18. Comparación de resultados entre los resultados de las Tablas 13 a 17.

De los resultados mostrados en la Tabla 18, se pueden hacer las siguientes apreciaciones:

- El puntaje global de consistencia de la sub-plantilla con mayor similitud a la plantilla, en el último ciclo de diezclado (factor 1) tiene valores entre aproximadamente 0.4 y 0.5, a excepción de la detección del movimiento idéntico que tiene un valor de aproximadamente 1. Estos puntajes siempre son mayores a los puntajes dados por otras sub-plantillas que no corresponden al movimiento descrito en *T* (Ver Anexo 2).
- Aunque el porcentaje de parches rojos y amarillos es menor para estos movimientos similares a la plantilla, con respecto a los valores mostrados en la Tabla 9 de las pruebas de calibración, se aprecia que estos porcentajes se encuentran entre valores similares para las 5 pruebas y son siempre mayores que los porcentajes de este mismo tipo de parches en los resultados provenientes de comparar una sub-plantilla con una plantilla, ambas con distintos movimientos (Ver Anexo 2).

Sin embargo, aunque los resultados fueron bastante favorables, se debe considerar que las pruebas generales fueron realizadas con videos que cumplían con ciertas características específicas necesarias para un correcto funcionamiento del sistema. Entre estas características, se encuentran las siguientes:

- Las dimensiones espacio-temporales son similares a las de los videos usados en las pruebas de calibración.
- Los objetos que realizan el movimiento de interés poseen dimensiones muy similares a las dimensiones del objeto que describe el movimiento en la plantilla.
- La plantilla seleccionada tiene, también, dimensiones espacio-tiempo muy similares a la plantilla utilizada en las pruebas de calibración.
- El movimiento de interés es muy específico y muy distinto a otros movimientos contenidos en los videos.
- Las dimensiones de los parches espacio-temporales son las mismas que las usadas en las pruebas de calibración.

6. CONCLUSIONES

El algoritmo desarrollado en este trabajo de grado tiene como objetivo detectar un movimiento específico en una secuencia de video. En este orden de ideas, el sistema desarrollado logra su objetivo con un porcentaje bastante alto de acierto, de acuerdo a los resultados obtenidos en las pruebas realizadas. Sin embargo, las pruebas generales fueron realizadas bajo un control riguroso de todos los factores que podrían afectar los resultados generados, y por tanto es posible que el algoritmo implementado no funcione tan bien en otro tipo de videos, con otro tipo de movimientos, y con valores diferentes en sus parámetros (a saber: dimensión de los parches, factor de diezmado, vecindad, desplazamiento entre sub-plantillas, etc.).

Por otro lado, debido al elevado costo computacional que tiene el procesamiento del algoritmo implementado, las pruebas que se realizaron (Calibración y Generales) fueron pensadas y seleccionadas con mucho rigor para no desaprovechar y malgastar el tiempo en una prueba infructuosa. A manera de ejemplo, una secuencia de video con una duración de aproximadamente 5 segundos, requería un tiempo de procesamiento de aproximadamente 1 hora y media. Además, como ocurrió en varias ocasiones durante las pruebas, cuando el video a analizar era un poco más largo, el equipo no podía realizar el procesamiento requerido y por tanto, se producía un error por falta de memoria virtual. Esto puede ser solucionado con un equipo con un procesador mucho más potente y una memoria RAM más grande; pero además con una depuración del código implementado en C++, haciendo el sistema más robusto y a la vez más eficiente.

Este algoritmo definitivo, es producto de un análisis detallado del algoritmo propuesto en [5], su posterior implementación en C++ y la ejecución de algunas modificaciones realizadas por parte de los autores de este trabajo de grado. Entre estas modificaciones se encuentran los bloques de detección movimiento, detección de un puntaje mayor de correlación entre sub-plantillas, y creación del video resultante. Todas estas modificaciones, explicadas durante el desarrollo del libro, surgen como respuesta a resultados indeseados producidos por el algoritmo original al ser implementado en C++.

Asimismo, cabe resaltar que a partir de los resultados obtenidos en las pruebas de calibración que, a diferencia de las pruebas generales, fueron realizadas con un video que contenía movimientos muy complejos y relativamente similares entre todos ellos, se llegó al desarrollo del

algoritmo definitivo. Las modificaciones consecuentes a las pruebas de calibración, establecieron la mejor combinación de parámetros (probada hasta ahora) que llevó a la obtención de resultados muy acertados. A saber, dicha combinación de parámetros es como sigue:

- El desplazamiento temporal fue definido fotograma a fotograma.
- El desplazamiento espacial fue definido pixel a pixel.
- El factor de diezmado fue definido como 3. Es decir, que se diezma dos veces el video original, realizando una búsqueda en el volumen total de un video con una resolución 4 veces menor a la del video original; luego se realiza una búsqueda en un video con una resolución 2 veces menor a la del video original en una vecindad alrededor de las coordenadas de la sub-plantilla con mayor puntaje; y finalmente, en el último ciclo de diezmado, se realiza una búsqueda en el video original en una vecindad alrededor de las coordenadas seleccionadas.
- La vecindad fue seleccionada como 3, logrando así una combinación entre un buen tiempo de procesamiento y una búsqueda con un buen nivel de exhaustividad.
- El factor de diezmado espacial fue definido como 2, siendo éste 1 vez menor que el diezmado espacial, de acuerdo a los resultados encontrados en la Sección 4.1.4.
- Los parches espacio-temporales fueron seleccionados con unas dimensiones de 7 píxeles de alto y 7 píxeles de ancho, logrando abarcar una región relativamente grande de movimiento pero sin producir un muy alto tiempo de procesamiento.
- La dimensión temporal de un parche fue definida en 3 fotogramas, que permite detectar cambios de movimiento a lo largo de cada parche sin generar un muy alto tiempo de proceso.

Como trabajo futuro, se propone realizar una modificación al algoritmo que permita detectar más de un movimiento. Esto es, para videos en los cuales existe no sólo 1 movimiento similar al descrito por la plantilla, sino varios. En principio, después de realizar la búsqueda a lo largo de todo el video en el primer ciclo de diezmado, se intentó realizar una selección de, no sólo uno sino, varios picos de consistencia global. Sin embargo esto implicaba que en el siguiente nivel de diezmado, la búsqueda no iba a ser sólo alrededor de la sub-plantilla con mayor puntaje global,

sino alrededor de varias sub-plantillas, lo cual a su vez implicaba una multiplicación del tiempo de procesamiento de acuerdo a la cantidad de picos encontrados. Es decir, si se quieren encontrar K movimientos, implica que el tiempo de procesamiento va a ser K veces mayor al tiempo de procesamiento actual del sistema. Además, otro problema adicional es la selección de un umbral adecuado para indicar que un pico de correlación efectivamente corresponde a un movimiento similar al descrito por la plantilla. Por esto, y obviamente por motivos de complejidad, se decidió detectar solamente el movimiento con mayor puntaje de consistencia global, esto es, la sub-plantilla que describe el movimiento más parecido al de la plantilla.

7. BIBLIOGRAFÍA

- [1] **A.A. Efros, A.C. Berg, G. Mori, and J. Malik**, “Recognizing Action at a Distance,” Proc. Int’l Conf. Computer Vision, Oct. 2003.
- [2] **A. Bobick and J. Davis**, “The Recognition of Human Movement Using Temporal Templates,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 23, no. 3, pp. 257-267, Mar. 2001.
- [3] **C. Bregler**, “Learning and Recognizing Human Dynamics in Video Sequences,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, June 1997, San Juan de Puerto Rico.
- [4] **Y. Yacoob and M.J. Black**, “Parametrized Modeling and Recognition of Activities,” IEEE Trans. Computer Vision and Image Understanding, vol. 73, no. 2, pp. 232-247, 1999.
- [5] **Schechtman Eli and Irani Michal**, “Space-Time Behavior-Based Correlation - OR - How to Tell If Two Underlying Motion Fields Are Similar without Computing Them?” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 29, no. 11, pp 2045-2056, Nov. 2007.
- [6] **Bovik AI**, *Handbook of Image and video processing*, Academic Press, 2000.
- [7] **Gary Bradski E Adrian Kaebler**, *Learning OpenCV: Computer Vision with the OpenCV Library*, O’Reilly, Software That Sees.
- [8] **G. Golub and C.V. Loan**, *Matrix Computations*, third ed. Johns Hopkins Univ. Press, 1996.

8. ANEXOS

8.1 Anexo 1: Código en C++

8.2 Anexo 2: Carpeta de resultados

8.3 Anexo 3: Artículo “Space-Time Behavior-Based Correlation” por Eli Shechtman y Michal Irani