ISSN 1870-4069

Mejoras al algoritmo de trayectorias densas para el reconocimiento de acciones en video

Fernando Camarena, Leonardo Chang, Miguel Gonzalez-Mendoza

Tecnológico de Monterrey, Campus Estado de México,
México

{a01370614, lchang, mgonza}@itesm.mx

Resumen. La habilidad para detectar personas y sus acciones de una manera autónoma y eficiente es uno de los objetivos principales de los sistemas inteligentes de video protección. El reconocimiento de acciones es parte importante de ello y en este trabajo exploramos diversas alternativas para mejorar el tiempo de ejecución y exactitud en uno de los métodos más usados: las trayectorias densas. Proponemos sustituir el algoritmo de flujo óptico Farneback por DisOF que permite reducir el tiempo de extracción de trayectorias en un 50%. De igual manera, analizamos la reducción del ruido provocado por las trayectorias no asociadas al objeto de interés mediante la estimación de los puntos anatómicos del cuerpo humano, discriminando más de la mitad de las trayectorias sin sacrificar de manera significativa la exactitud de los resultados. Adicional a esto, exploramos la idea de incorporar las relaciones espaciales entre trayectorias a través del uso de la técnica de pirámide espacial, encontrando que es posible mejorar la eficacia en los resultados.

Palabras clave: trayectorias densas, reconocimiento de acciones, visión por computadora, estimación de postura, relaciones espaciales.

Improvements to the Dense Trajectories Algorithm for Action Recognition

Abstract. The ability to detect people and their actions in an autonomous and efficient way is one of the main objectives of intelligent video-protection systems. Action recognition is one of the most important parts of this kind of systems. In this work, we explore diverse alternatives to improve both the accuracy and execution time in one of the most used methods: dense trajectories. We propose to replace the optical flow algorithm from Farneback to DisOF, our results show that the time needed to extract the dense trajectories is reduced by 50%. Also, we analyze how the noisy trajectories can be reduced by estimating the anatomical points of the human body. In this way, more than half of the total trajectories were eliminated without a significant loss of accuracy. In addition to this, we study how spatial relationships through the spatial pyramid technique

can be applied to the dense trajectories method, resulting in an improvement to the accuracy.

Keywords: dense trajectories, action recognition, computer vision, pose estimation, spatial relation.

1. Introducción

La seguridad e integridad de las personas es un problema que todo gobierno, industria e instituciones académicas enfrentan. Los sistemas de video protección se han convertido en uno de los medios más populares debido a su accesibilidad para los usuarios. Actualmente, técnicas de visión por computador y aprendizaje de máquina [12] han buscado romper con uno de los mayores problemas de los sistemas actuales; su dependencia con la intervención humana.

El reconocimiento de acciones es una de las áreas más importantes que forman la video protección. Recientemente, su estudio se encuentra enfocado en la extracción y clasificación de trayectorias densas [1,2]. La idea principal del método consiste en realizar un muestreo denso sobre la imagen, seguir cada uno de estos puntos a lo largo de los cuadros que conforman el video y posteriormente describir esta secuencia o flujo tanto en su forma, movimiento y apariencia local.

El proceso de extracción de trayectorias generará una cantidad indefinida de ellas, por lo que es necesario aplicar un método que asegure tener una salida estándar para cada secuencia de video. Comúnmente se suele aplicar un enfoque de bolsa de palabras, que consiste en la formación grupos o palabras visuales mediante la identificación de trayectorias similares [8]. El resultado será un histograma de ocurrencias de las palabras visuales.

A pesar de ser un método que ha permitido tener excelentes resultados [8], su naturaleza de agrupar y entregar un histograma de ocurrencias hace que se pierda una de las características más importantes para el reconocimiento de acciones: la relación. En este contexto significaría que no podríamos saber si el movimiento descrito proviene de una mano o una pierna, por citar un ejemplo. Contar con esta información puede llegar a representar una mejora significativa.

Otro aspecto clave en los sistemas de video protección es su capacidad de procesar las imágenes en tiempo real, por lo que tener algoritmos cada vez más eficientes se traduciría en poder llevar estas técnicas a un contexto de aplicación.

En este trabajo buscamos explorar cada uno de los problemas mencionados; a través de la estimación de los puntos anatómicos es posible realizar una segmentación del cuerpo humano y en conjunto a la técnica de pirámide espacial añadir este tipo de relación. De igual forma, podemos utilizar los puntos anatómicos para filtrar las trayectorias que pertenecen al fondo del video y así disminuir la complejidad y ruido. Nuestra experimentación muestra que al tomar en cuenta las relaciones espaciales la exactitud se ve beneficiada y que el filtrado de trayectorias permite disminuir los tiempos de ejecución sin sacrificar de manera significativa la exactitud.

Fig. 1. Proceso de extracción de trayectorias. El proceso de extracción de trayectorias está dividido en tres fases: Muestreo, seguimiento y descripción. En la fase de muestreo se generan una serie de puntos de interés en diferentes escalas de la imagen. La segunda fase tomará cada escala de manera independiente y realizará un seguimiento de los puntos de interés a lo largo de L cantidad de cuadros. Por último, la trayectoria formada por el seguimiento de los puntos debe de ser descrita tanto en su forma, movimiento y apariencia local por medio de los descriptores HOG, HOF y MBH.

Por otra parte, el cálculo de flujo óptico mediante la búsqueda inversa densa (del inglés, Dense Inverse Search, DisOF) [3] ha superado a los algoritmos actuales tanto en calidad como rapidez. Nuestros resultados muestran que su uso disminuye significativamente el tiempo necesario para la extracción de trayectorias densas al mismo tiempo que la exactitud de los resultados mejora.

Este trabajo está organizado de la siguiente manera: la sección 2 describe el proceso de clasificación de acciones mediante el uso de trayectorias densas. En la sección 3 describimos a detalle nuestra propuesta para mejorar la exactitud y tiempo de ejecución del proceso. En la sección 4 describimos los experimentos y conjunto de datos utilizados. En la sección 5 presentaremos y discutiremos los resultados obtenidos para dar paso a las conclusiones en la sección 6. Por último, presentaremos el futuro de nuestro trabajo en la sección 7.

2. Reconocimiento de acciones mediante el uso de trayectorias densas

El uso de trayectorias densas ha mostrado ser un medio efectivo para representar videos. Heng y colaboradores [1] proponen un método para la extracción de trayectorias y su clasificación. Este proceso se divide en tres fases: extracción de trayectorias (ver figura 1), obtención de descriptores mediante una bolsa de palabras (ver figura 2) y la clasificación.

2.1. Extracción de trayectorias

El proceso de extracción de trayectorias se compone de tres fases principales: muestreo denso, seguimiento de puntos y descripción de la secuencia. El proceso de muestreo consiste en generar una serie de puntos a lo largo de diferentes escalas de un cuadro del video.

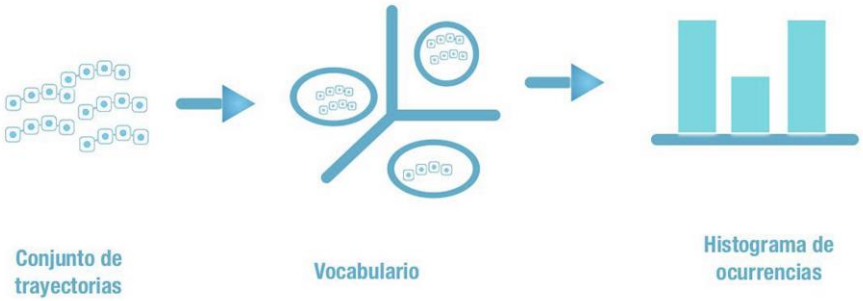

Fig. 2. Ejemplo de método basado en bolsas de palabras. El primer paso consiste en generar un conjunto de palabras representativas (vocabulario) por medio de la identificación y agrupamiento de las trayectorias similares en el conjunto de entrenamiento. El segundo paso consiste en asociar las trayectorias de una secuencia de video a su correspondiente palabra representativa y el descriptor será un histograma de ocurrencias de cada una de ellas.

El siguiente paso consiste en realizar el seguimiento de cada punto, recordando que las escalas se manejan de manera independiente. Para llevarlo a acabo es necesario utilizar algún método de flujo óptico; Heng y colaboradores [1,2] proponen utilizar Farneback [5] debido a su balance entre desempeño y eficiencia.

El último paso consiste en describir el flujo obtenido, para ello se define una ventana que contenga información de la vecindad. De esta forma será posible describir la trayectoria según su forma, movimiento y apariencia local por medio de los descriptores HOG, HOF y MBH [6,7].

2.2. Bolsa de palabras

Cada secuencia tendrá asociada una cantidad variable de trayectorias, por lo que es necesario aplicar un método que permita estandarizar la salida, de tal forma que pueda ser utilizada por un clasificador. Los métodos basados en bolsas de palabras (del inglés, Bag of Words, BoW) han sido de los enfoques más utilizados debido a sus resultados destacables en los años recientes [8,9]. La idea principal consiste en realizar un agrupamiento de características (trayectorias) pertenecientes al conjunto de entrenamiento para identificar aquellas que sean similares; cada uno de estos grupos representará una palabra visual y su conjunto formará lo que se conoce como vocabulario visual. Posteriormente cada trayectoria de un video debe de ser asociada con alguna palabra visual del vocabulario y, por tanto, el descriptor será un histograma de las ocurrencias de cada palabra visual (ver figura 2).

Una de las principales limitaciones de estos enfoques es que el vocabulario visual se construye utilizando las trayectorias que pertenecen al objeto y al fondo; esto implica que el ruido que existe en la imagen será considerado como un objeto de la clase. De igual manera al tratarse de un histograma de ocurrencias cualquier relación espacial entre las trayectorias se pierde, característica que contiene información útil para

identificar una acción. Perder la relación espacial implicaría desconocer de qué parte del cuerpo proviene la trayectoria y cómo esta se relaciona con las demás.

2.3. Clasificación

El último paso en el proceso para reconocer una acción por medio de trayectorias densas es la clasificación. Uno de los clasificadores más utilizados son las máquinas de soporte de vectores (SVM), en este caso se utiliza un SVM no lineal con un kernel χ^2.

3. Propuesta

En la sección 2, identificamos que la naturaleza de los métodos basados en bolsas de palabras tiene como limitante que suelen incluir ruido como parte de las palabras visuales y además su funcionamiento basado en ocurrencias elimina la posibilidad de incluir relaciones espaciales. De igual forma, observamos que existe una necesidad por reducir los tiempos de ejecución de los algoritmos.

En este orden de ideas, proponemos explorar el uso y estimación de los puntos anatómicos para la segmentación del cuerpo humano (ver Figura 3). Esto, indudablemente, permite atacar el problema de incluir ruido producido por el movimiento de fondo al conocer qué trayectorias pertenecen al sujeto. Por otra parte, uno de los métodos que ha mostrado ser efectivo, en otras áreas, para incorporar relaciones espaciales es el uso de la técnica de pirámide espacial [10], que consiste en describir regiones uniformes de la imagen y posteriormente utilizar como descriptor su concatenación. En el presente trabajo, proponemos aplicar este enfoque utilizando como regiones la segmentación obtenida por los puntos anatómicos.

Para la estimación de estos puntos, utilizaremos el enfoque de Cao y colaboradores [4], cuyo método ha superado en eficiencia y exactitud a los algoritmos presentados en el MPII Multi-Person Benchmark y Coco 2016, keypoints challenge.

Por otro lado, el cálculo de flujo óptico es una pieza fundamental en el proceso de extracción de las trayectorias densas y una de las técnicas más utilizadas en el área de visión por computador, por lo que su investigación se encuentra en mira. Kroeger y colaboradores [3] presentan un nuevo método que se basa en la búsqueda densa inversa y supera tanto en eficiencia como en exactitud al resto de los algoritmos reportados en la literatura. Por tanto, proponemos como parte de este trabajo, utilizar el algoritmo DisOF para el cálculo de las trayectorias densas, con lo que esperamos obtener una mejora tanto en los tiempos de ejecución, como exactitud de los resultados.

Esto nos lleva a modificar el proceso de reconocimiento de acciones mediante el uso de trayectorias densas de la siguiente manera: El proceso de extracción de trayectorias cambia su algoritmo de flujo óptico por DisOF. Adicionamos una nueva capa (ver figura 4) entre extracción de trayectorias y creación de vocabulario que se encarga de identificar y clasificar las trayectorias de acuerdo con los puntos anatómicos del cuerpo. La generación de vocabulario y obtención de descriptores se hará para la imagen completa con sus partes.

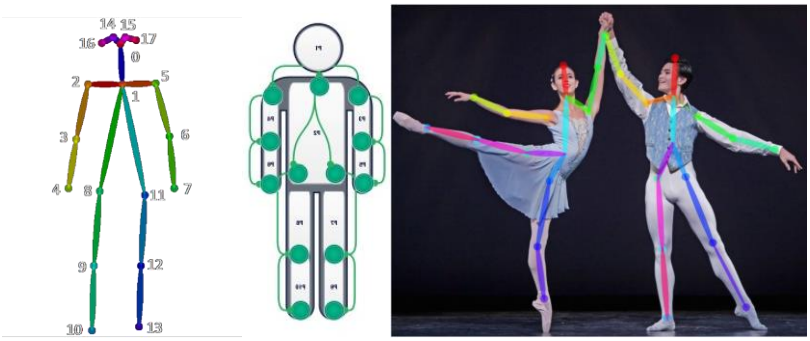

Fig. 3. A la izquierda podemos visualizar los puntos anatómicos generados por Cao y colaboradores [4], la imagen central indica cómo es posible segmentar el cuerpo por medio de estos puntos y la imagen de la derecha podemos ver cómo la relación espacial entre estos puntos es puede ser un indicativo a la acción que se está realizando.

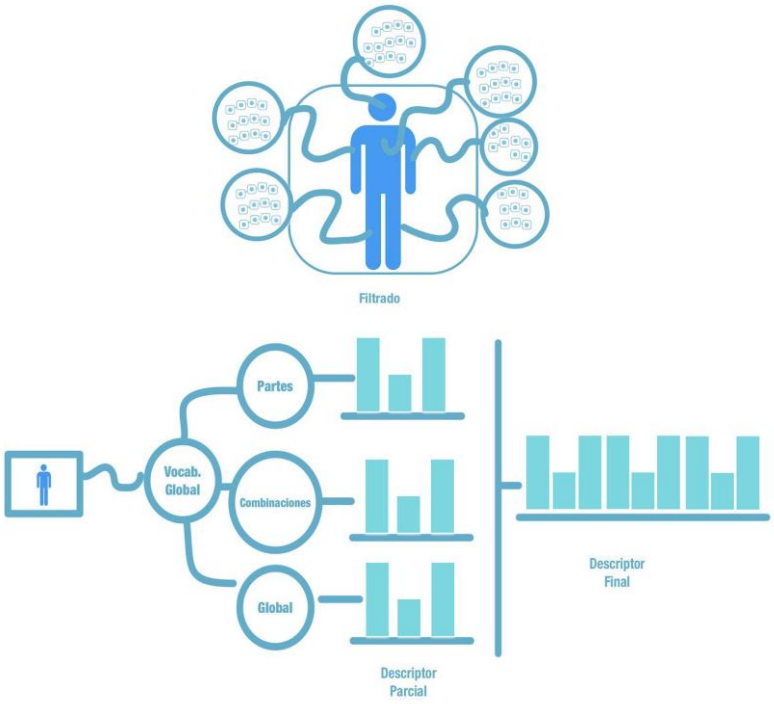

Fig. 4. Nuestra propuesta para mejorar el proceso de reconocimiento de acciones mediante el uso de trayectorias densas. Se añadió un nuevo módulo encargado de asociar las trayectorias con alguna parte del cuerpo. Posteriormente, cada una de las partes y combinaciones entre ellas serán utilizadas para obtener descriptores parciales, que darán vida a un únicom descriptor mediante su concatenación. Para acelerar el proceso y mejorar la exactitud proponemos utilizar (DISOF) como elemento de flujo óptico en la extracción de trayectorias.

4. Configuración de los experimentos

En esta sección describiremos el conjunto de datos utilizado en los experimentos, los parámetros utilizados en los algoritmos y las diferentes configuraciones de los experimentos. Las pruebas se realizaron utilizando una computadora personal con Macos High Sierra que cuenta con un procesador Intel Core i7 (séptima generación a 2.9GHz) con 16 GB en RAM.

4.1. KTH dataset

Utilizamos el conjunto de datos KTH [11] que se conforma de 6 diferentes de acciones: Caminar, Trotar, correr, boxear, saludar y aplaudir. Cada una ellas son ejecutadas varias veces por 25 sujetos en 4 diferentes escenarios (interior, exterior, exterior con variación de escala, exterior con diferentes tipos de ropa). En total se cuentan con 2391 secuencias, mismas que fueron grabadas con fondo homogéneo por medio de una cámara estática a 25 cuadros por segundo.

Los experimentos fueron realizados utilizando configuración descrita por los autores. El conjunto de pruebas está formado por los sujetos 2, 3, 5, 6, 7, 8, 9, 10 y 22, mientras que el resto forma el de entrenamiento. Como medida de desempeño utilizamos la exactitud obtenida en el conjunto de pruebas.

4.2. Trayectorias densas

El proceso para la extracción de trayectorias se hará utilizando el código fuente del autor con sus parámetros por defectos [1].

Para el algoritmo de flujo óptico DisOF utilizaremos la implementación disponible en OpenCV. En el caso de la generación de vocabulario y descriptores seguiremos los lineamientos descritos en [8,9] utilizando la configuración descrita por Heng y colaboradores [1]: 4000 palabras visuales y un vocabulario por descriptor. Para la clasificación utilizaremos un SVM no lineal con un kernel χ^2 disponible en la librería de OPENCV.

4.3. Estimación de puntos anatómicos

Para estimar los puntos anatómicos utilizaremos el método propuesto por Cao y colaboradores [4] y posteriormente utilizaremos estos puntos para generar las regiones de cada parte del cuerpo. En total generamos 10 regiones (Torso, cabeza, área de los bíceps, área de los antebrazos, área de los cuádriceps y área de los gemelos) (ver figura 3).

4.4. Descripción de los experimentos

Presentamos dos grupos de experimentos; el primero de ellos consiste en atender la necesidad de mejorar el tiempo de ejecución de los algoritmos de extracción y

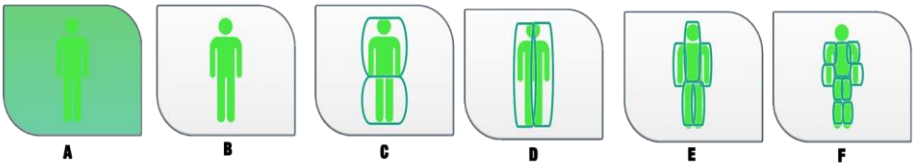

Fig. 5. Diferentes agrupamientos de las partes del cuerpo para probar las relaciones espaciales. Cada para de la configuración añade cierto nivel de relación espacial. Combinar varias de estas configuraciones puede resultar un buen enfoque para añadir la característica a este proceso.

Tabla 1. Descripción del algoritmo para determinar a qué parte del cuerpo pertenece una trayectoria. Su funcionamiento se basa en identificar si el punto denso muestreado en la primera fase se encuentra dentro de alguna región de interés; en caso de existir oclusión incorporamos un sistema de votación que se encarga de determinar a qué parte del cuerpo pertenece la trayectoria.

#	**Descripción**
1	Del descriptor de la trayectoria tomar el punto (X, Y), que es la media de todos los puntos por los que paso la trayectoria.
2	Del descriptor de la trayectoria tomar la información sobre el cuadro de terminación y la longitud de la trayectoria. Esto permite conocer el cuadro de inicio (*CuadroInicio = cuadro_de_terminación – longitud).*
3	Inicializar un arreglo de tamaño 11 (10 partes + 1 fondo) donde guardaremos los votos.
4	Para cada uno de los cuadros de la trayectoria verificar si *Media* se encuentra en alguna región del cuerpo y sumar 1 a la posición del arreglo correspondiente.
5	La trayectoria pertenece a la parte con el mayor número de votos.

clasificación de las trayectorias. Nuestro segundo grupo de experimentos busca mejorar la exactitud de los algoritmos mediante la exploración de las relaciones espaciales.

Para conocer el efecto de DisOF [3] como algoritmo de flujo óptico en la extracción de trayectorias usaremos dos parámetros: la exactitud y el tiempo necesario para procesar cada cuadro. La exactitud será la relación entre los elementos bien clasificados frente al total de elementos. La ecuación (1) muestra el tiempo de ejecución, que estará dado por la relación entre el tiempo y la cantidad de cuadros por secuencia y para añadir mayor confianza tomaremos el promedio de haber analizado todas las secuencias contenidas en 100 videos.

$$Tc = Ts / N,$$
$$Tc = Tiempo\ por\ cuadro,\ Ts = Tiempo\_secuencia,\ N = número\ de\ cuadros. \quad (1)$$

Uno de los problemas identificados con el uso de bolsas de palabras fue la de incorporación de ruido a la hora de crear el vocabulario. Para solventarlo proponemos en la utilizar el método de Cao y colaboradores [4] para generar los puntos anatómicos del cuerpo y posteriormente dividir cada una de las trayectorias en alguna de las 11

Tabla 2. Resultados de comparar el proceso de Heng y colaboradores [1,2] con nuestra propuesta (A + DisOF [3]). Los resultados muestran que nuestro método supera tanto en rapidez como en exactitud al método propuesto por Heng y colaboradores.

Configuración	Tiempo (segundos)	Exactitud (%)
Heng y colaboradores [1,2]	0.0684152	95.83
Nuestra propuesta (A + DisOF)	**0.03899**	**96.29**

Tabla 3. Comparación de los tiempos de ejecución para la creación del vocabulario final y descriptores con y sin nuestro método de filtrado de trayectorias. Los resultados muestran que el filtrado de trayectorias suele ofrecer una mejoría pequeña en el tiempo de ejecución para la mayoría de los descriptores.

	Trayectoria (s)	HOG (s)	HOF (s)	MBHX (s)	MBHY (s)
Construcción de vocabulario visual sin nuestro método de filtrado.	489.90	523.01	**410.58**	**418.05**	473.71
Construcción de vocabulario visual con nuestro método de filtrado.	**339.84**	**469.34**	547.1	427.47	**473.07**
Obtención de descriptores sin nuestro método de filtrado			562.72		
Obtención de descriptores con nuestro método de filtrado			**222.05**		

categorías disponibles. (10 partes del cuerpo + fondo). La descripción detallada del proceso se encuentra descrito en la tabla 1.0.

Con la discriminación de las trayectorias pertenecientes al fondo pretendemos reducir el tiempo de procesamiento en fases posteriores, sobre todo en la creación del vocabulario visual e histograma de ocurrencias. Para medir qué tan eficiente resultado, simplemente tomaremos el tiempo de ejecución que nuestra máquina tarda en crear el vocabulario (por descriptor) y cuánto tiempo le lleva generar todos descriptores finales.

Nuestro segundo grupo de experimentos tiene como objetivo explorar el uso de relaciones espaciales mediante la técnica de pirámide espacial [10]. Para lograr este efecto utilizamos las partes del cuerpo humano en diferentes combinaciones (ver figura 5).

5. Resultados

El presente trabajo tiene dos principales objetivos: explorar mejoras a los algoritmos descritos para reducir el tiempo de ejecución y la incorporación de las relaciones espaciales.

Tabla 4. Comparación de los resultados después de haber aplicado haber tomado en cuenta las relaciones espaciales. Podemos notar que algunas configuraciones ayudan a mejorar la exactitud de los resultados.

Configuración	Exactitud (%)
1: Nuestra propuesta (A + C + F)	95.37
2: Nuestra propuesta (B + C +F)	94.90
3: Nuestra propuesta (B + C)	95.83
4: Nuestra propuesta (B + F)	94.907
5: Nuestra propuesta (A + C)	**96.75**
6: Nuestra propuesta (A +F B + C)	96.29
7: Nuestra propuesta (A + B + C + D)	**96.75**
8: Nuestra propuesta (A + B + C + D+ E)	96.29
9: Nuestra propuesta (F)	91.20
10: Heng y colaboradores [1,2]	95.83

5.1. Reducción del tiempo de ejecución

La tabla 2 muestra los resultados del primer experimento. La configuración descrita por Heng y colaboradores obtiene una exactitud de 95.83%, cuyo proceso de extracción de trayectorias toma un total de 0.0684152 segundos por cuadro. La incorporación del algoritmo DisOF [3] permite llegar a una exactitud de 96.29%, donde el proceso de extracción de características por cuadro es de 0.03899 segundos. Podemos concluir que su incorporación no sólo involucra un desempeño mejor, sino que el tiempo de ejecución del algoritmo prácticamente duplica la velocidad del algoritmo original de Farneback.

A continuación, se presentan los resultados del filtrado de trayectorias pertenecientes al fondo. Al utilizar el algoritmo descrito en la tabla 1 para asociar las trayectorias a una parte del cuerpo encontramos que de las 3,820,795 trayectorias de todo el conjunto de datos, 2,168,235 trayectorias pertenecen al fondo, lo que representa un 56.74% de ruido. Esto sin duda es una reducción importante que impacta en el tiempo de ejecución de las fases posteriores a la extracción. La tabla 3 muestra los resultados obtenidos al utilizar este subconjunto de trayectorias, como era de esperar el tiempo de ejecución disminuyó, pero el proceso provocó que la exactitud bajará a un 95.3%.

5.2. Exploración del uso de relaciones espaciales

Nuestro segundo grupo de experimento tiene la finalidad de explorar cómo afecta las relaciones espaciales mediante el uso de la técnica pirámide espacial al proceso de clasificación de acciones por trayectorias densas. Para ello utilizamos la concatenación de los descriptores después aplicado a diferentes grupos de partes del cuerpo. La tabla 4.0 muestra los resultados de cada una de las configuraciones y podemos apreciar que las relaciones espaciales mediante el uso de una de pirámide espacial no es la mejor

forma para describir una acción. Sin embargo, podemos notar que en ciertos casos lograron mejorar la exactitud del clasificador.

Por lo que podemos concluir que las relaciones espaciales pueden ayudar a mejorar un clasificador de acciones. Sin embargo, aplicarlo utilizando una pirámide espacial no resulta ser la opción más adecuada para describir acciones.

6. Conclusiones

El uso de trayectorias densas ha mostrado tener un buen desempeño para describir video, pero encontramos que su construcción cuenta con ciertas áreas de mejora. El primero de ellos consiste en la necesidad de contar con algoritmos que funcionen en tiempo real; para solucionar proponemos de DisOF [3] como algoritmo de flujo óptico, encontramos que los resultados permitieron reducir casi a la mitad los tiempos de ejecución para la extracción de las trayectorias además que la exactitud del clasificador se ve beneficiada.

Exploramos tratar algunas deficiencias del uso de métodos basados en bolsas de palabras, mediante el uso de Cao y colaboradores [4] filtramos las trayectorias por del fondo del video, este método ayudó a reducir de manera significativa la cantidad de trayectorias, mejorando el tiempo de ejecución en fases posteriores a la extracción sin comprometer de manera significativa la exactitud de los resultados.

Probamos añadir relaciones espaciales mediante la técnica de pirámide espacial. Encontramos que el método es capaz de mejorar la exactitud de los resultados e indica que la investigación en técnicas más adecuadas es un camino que seguir.

7. Trabajo a futuro

Una de las conclusiones más interesantes de este trabajo es encontrar que las relaciones espaciales ayudan a mejorar la exactitud de la clasificación, basado en este principio nos enfocaremos en estudiar y explorar diferentes formas de incorporar estas relaciones. Contamos con la información de los puntos anatómicos del cuerpo, por lo que crear un descriptor con ellos puede resultar una mejor manera de representar la relación. Otro aspecto de mejora está nuestro algoritmo para dividir las trayectorias, debido a que utilizamos un método trivial para manejar las oclusiones, por lo que formas más complejas podrían ayudar a incrementar la exactitud en los datos.

De igual manera, el trabajo presentado se basa en un único conjunto de datos, por lo que procederemos en extender estos métodos a conjuntos de datos más grandes y complejos.

Referencias

1. Wang, H., Kläser, A., Schmid, C., Liu, C.L.: Action recognition by dense trajectories. In: Computer Vision and Pattern Recognition (CVPR), pp. 3169–3176 (2011)

2. Kroeger, T., Timofte, R., Dai, D., Van Gool, L.: Fast optical flow using dense inverse search. In: European Conference on Computer Vision, pp. 471–488 (2016)
3. Cao, Z., Simon, T., Wei, S.E., Sheikh, Y.; Realtime multi-person 2d pose estimation using part affinity fields. In: CVPR 1(2), pp. 7 (2017)
4. Farnebäck, G.: Two-frame motion estimation based on polynomial expansion. In: Scandinavian conference on Image analysis, pp. 363–370 (2003)
5. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Computer Vision and Pattern Recognition, (CVPR), IEEE Computer Society Conference on 1, pp. 886–893 (2005)
6. Dalal, N., Triggs, B., Schmid, C.: Human detection using oriented histograms of flow and appearance. In: European conference on computer vision, pp. 428–441 (2006)
7. Chang, L., Pérez-Suárez, A., Hernández-Palancar, J., Arias-Estrada, M., Sucar, L.E.: Improving visual vocabularies: a more discriminative, representative and compact bag of visual words. Informática 41(3) (2017)
8. Chang, L., Pérez-Suárez, A., Rodríguez-Collada, M., Hernández-Palancar, J., Arias-Estrada, M., Sucar, L.E.: Assessing the Distinctiveness and Representativeness of Visual Vocabularies. In: Iberoamerican Congress on Pattern Recognition, pp. 331–338 (2015)
9. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: Computer vision and pattern recognition, IEEE computer society conference on 2, pp. 2169–2178 (2006)
10. Schuldt, C., Laptev, I., Caputo, B.: Recognizing human actions: a local SVM approach. In: Pattern Recognition, (ICPR´04). In: Proceedings of the 17th International Conference on 3, pp. 32–36 (2004)
11. Borges, P. V. K., Conci, N., Cavallaro, A.: Video-based human behavior understanding: A survey. IEEE transactions on circuits and systems for video technology 23(11), pp. 1993–2008 (2013)