Sustracción de fondo por varias características estables en el modelo

Leonardo Dominguez[1,2], Alejandro Perez[1,3], Juan P. D'Amato[1,2] y Rosana Barbuzza[1,3]

[1] PLADEMA, Universidad Nacional del Centro de la Provincia de Buenos Aires,
[2] Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET
[3] Comisión de Investigaciones Científicas, CICPBA

Resumen. Los métodos de sustracción de fondo basados en modelo con una única característica como la intensidad del píxel, suelen fallar en la clasificación de escenas complejas. En este trabajo se propone ampliar los descriptores del modelo de fondo para considerar otras características como la textura, la distribución de intensidades, escala de grises, color, y de esta manera mejorar la clasificación de cada píxel. Para clasificar además se tiene en cuenta la característica principal y secundaria en cada región de imágenes tomadas con cámaras estáticas. En particular para la textura, se utilizó una modificación del descriptor simple y tradicional Local Binary Pattern (LBP) que resulta invariante a los cambios de tonalidades en escala de grises, y la rotación. Los descriptores fueron incorporados al algoritmo de sustracción de fondo *Visual Background Extraction (ViBE)*, que identifica zonas de movimiento en las escenas, comparando distintas características del modelo de fondo. El algoritmo propuesto se puede aplicar para detectar personas o vehículos en aplicaciones para seguridad ciudadana, monitoreo de tráfico, entre otros. Los resultados preliminares obtenidos en la detección de objetos muestran que es factible utilizar varios descriptores del modelo de fondo para lograr mejorar la tasa de acierto y con bajo costo computacional, con la consiguiente ventaja para etapas de procesamiento posteriores, como el reconocimiento y el seguimiento de los objetos.

Palabras clave: descriptores de textura, patrones invariantes, detección de objetos

1 Introducción

La sustracción de fondo tiene un rango de aplicaciones muy amplio, que incluye para el área de seguridad, el monitoreo a través de video. En los últimos años ha crecido la cantidad de centros de monitoreo, con gran cantidad de operadores observando cámaras para detectar situaciones de interés como vandalismo, infracciones de tránsito, etc [1]. En esta línea, este grupo de investigación ha presentado varios trabajos, principalmente mostrando la arquitectura de un sistema abierto, distribuido y escalable [2] [3][4].

Con los recientes avances en modelos de fondo, la sustracción se ha vuelto más práctica y atractiva para el área de visualización computacional. Tradicionalmente, un píxel se rotula como *background* cuando la variación de características está bajo un umbral de estabilidad en la escena. El principal desafío en la sustracción de fondo viene de los ambientes dinámicos y complejos, tales como árboles en movimiento, luminosidad del ambiente, etc. Otro de los problemas de la sustracción de fondo es descartar la sombra de los objetos en movimiento, ya que afecta considerablemente el tamaño, color u orientación del mismo que afecta drásticamente un posterior análisis del objeto para su identificación, clasificación o seguimiento [5]. En este trabajo, se detectan zonas de movimiento en la escena, y se utilizan indicadores para comparar los resultados obtenidos con métodos del estado del arte. Con la incorporación de textura en el modelo de varias características se busca separar también las sombras, mejorando la forma del objeto detectado. El trabajo se organiza de la siguiente manera. En la sección 2 se detalla el estado del arte, y en la sección 3 el método de sustracción de fondo ViBE y las modificaciones realizadas. En las secciones 4 y 5 la clasificación por texturas y de resultados obtenidos. Finalmente en la sección 6 se presentan las conclusiones y futuros trabajos.

2 Estado del arte

Un píxel se describe mejor utilizando varias características visuales como el color, escala de grises, textura, la distribución de intensidades, entre otros. En general, existe una característica que es dominante y puede ser más estable que el resto en cada región de la imagen. Las características del fondo de la escena siempre son bastante invariantes o cambian poco a través de una secuencia, por ejemplo la iluminación a través del día.

Los algoritmos clásicos utilizan una sola característica para modelar el fondo. Algunos de estos métodos como *Gaussian* o *mixture of Gaussians* (MoG) [5], así también como los métodos estocásticos *ViBE* o *ViBE+* [6] [7], se basan en información de color de los píxeles. Otros métodos utilizan la característica de textura para modelar el fondo, como los propuestos inicialmente en [8] [9] y extensiones como en [10]. Otra categoría de algoritmos se basan en alguna característica temporal (*interframe*) para detectar objetos en movimiento y separarlos del fondo [11].

Existen otros algoritmos que se basan en multicaracterísticas para modelar el fondo. Cada característica tiene su propia desventaja, por ejemplo considerar escala de grises es muy sensitivo a los cambios de iluminación, mientras que la textura, no puede separar correctamente cuando el fondo y el objeto son similares. La integración de ambas características da una forma más efectiva de mejorar la tasa de aciertos, para lograr un sistema simple y robusto. En este trabajo, se eligen las características de escala de grises del píxel, las componentes de color RGB y el descriptor de textura RSILTP, que es una extensión del patrón binario tradicional LBP [12], y tiene mayor robustez a los cambios de iluminación. El descriptor RSILTP [9] requiere poco almacenamiento, y usa solo un bit para

codificar las diferencias entre el píxel central y cada vecino, siendo que otros descriptores como LBP codifican usando vectores de 8 bits. Tanto el valor de intensidad en grises como el descriptor modificado de RSILTP se agrega al método de sustracción de fondo ViBE [2] [7], a diferencia de otros métodos como [13] [14] que utilizan MoG como sustractor de fondo. En este trabajo se considera que la característica dominante es el color de la escena y la característica secundaria es la textura para clasificar. Sin embargo, a diferencia de estos últimos autores [13] [14], se trabaja primero procesando por color y luego sobre estos resultados se resuelve sobre la textura. De esta forma, el procesamiento por separado permite introducir el concepto de corregir aquellos píxeles que son vecinos al objeto detectado en movimiento por la característica de color de la primera etapa, y que es información para la segunda etapa del procesamiento por textura. Esto evita calcular el indicador de textura para clasificar píxeles que son invariantes en el tiempo. También, en forma diferente se compara cada píxel de la imagen con la textura del modelo de fondo y no en forma separada por textura de la imagen y textura del modelo del fondo como en [13]. Esto permite comparar las texturas que pueden ser uniformes en la imagen y en el fondo, pero que tienen diferencia en la escala de intensidad.

3 Sustractor de fondo en video

En este trabajo se utiliza el sustractor de fondo ViBE, propuesto en [7], extendiendo el modelo de la propuesta original de única característica (color) a multicaracterística (color, textura, etc). Entre las virtudes de este método ViBE se destacan el bajo tiempo de cómputo, las altas tasas de detección y la robustez ante la existencia de ruido, las cuales son necesarias en capturas de cámaras de supervisión utilizadas hoy en día. Igualmente, la propuesta puede llegar a aplicarse a otros algoritmos [5]. En la Figura 1, se muestra la sustracción de fondo realizada con ViBE para un *frame* particular para un video *Highway* [15]. En el ejemplo, se puede ver la imagen original y a su derecha, las componentes *foreground* detectadas de los autos, y el *frame groundtrue* de la base de datos.

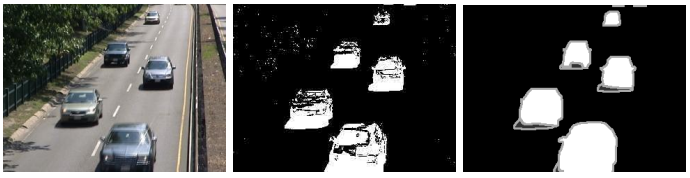

Fig. 1. Imagen del video (*izq.*) y sustracción del fondo con ViBE (*centro.*)[7] y *groundtrue* (*der.*), para un *frame* del video Highway

El método ViBE utiliza un modelo del fondo que almacena para cada píxel, N muestras seleccionadas aleatoriamente de *frames* ya procesados. Esto es un modelo basado en una única característica, el color. Si consideramos m=$\{m_1, m_2, ..., m_N\}$

las muestras correspondientes al píxel actual p_t, cada m_i para i=1..N, consiste en un vector m_i=(r,g,b) que usa los componentes del espacio RGB. Esto puede adaptarse a utilizar menor cantidad de componentes, escala de grises u otro. Luego, se define la condición de intersección I(p_t, m_i) igual a 1, cuando la distancia euclídea D, entre estos parámetros es menor o igual a un determinado valor de radio R, es decir:

$$I(p_t, m_i) = \begin{cases} 1, & \text{si } D(p_t, m_i) \leq R \\ 0, & \textit{en otro caso}, \end{cases} \quad (1)$$

La clasificación consiste en comparar cada píxel de la imagen actual con las N muestras del mismo que se encuentran almacenadas en el vector m, y mediante un indicador determinar si es *background* o *foreground*. Luego, la siguiente función,

$$Deteccion(p_t, m) = \begin{cases} 0, & \text{si } \sum_{i=1}^{N} I(p_t, m_i) \geq \#Min \\ 1, & \textit{en otro caso}, \end{cases} \quad (2)$$

define una máscara binaria de clasificación para cada píxel del *frame* de entrada, #Min es la cantidad mínima de veces que debe ser verdadera la Ec. 1 para considerarse *background*. En el artículo [7], se propone utilizar #Min=2, N=20 y R=20.

Como consecuencia, por cada imagen del video, se genera una salida binaria correspondiente a la clasificación de cada píxel. Luego de la detección, el método ViBE actualiza el modelo en forma estocástica, para adaptar gradualmente la representación a los diferentes cambios que ocurren a lo largo de una secuencia.

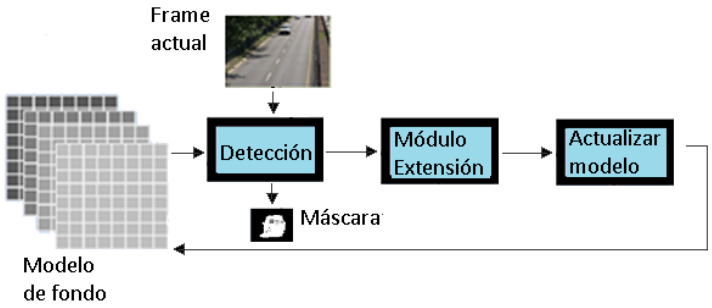

Fig. 2. Algoritmo ViBE y módulo extendido con mejoras implementadas

En la Figura 2, se muestra el método ViBE original y los dos módulos básicos Detección y Actualizar el modelo de fondo. La actualización del modelo de fondo para cada píxel es aleatoria, con cierta probabilidad de reemplazar uno de los valores guardados en el modelo por un nuevo valor de intensidad del mismo píxel en el *frame* actual. La aleatoriedad sobre este mecanismo de actualización permite reducir gradualmente con el tiempo, la probabilidad de que las muestras

guardadas persistan en el modelo [7]. Este modelo es representativo del fondo, por eso es deseable que sólo aquellos píxeles clasificados como *background* según la Ec. 2 se inserten en el modelo, ya que la inserción de píxeles que pertenecen a objetos en movimiento, o que resultan inciertos en la clasificación, pueden alterar significativamente los resultados de la detección de futuros píxeles. Esto último es muy importante ya que permite la introducción de mejoras al método ViBE logrando mayor tasa de detección. Con el módulo de extensión se realiza la corrección de píxeles mal clasificados antes de actualizar el modelo, lo que permite que el modelo de fondo sea confiable. El módulo de extensión propone corregir los píxeles clasificados como *background* pero que se corresponden a *foreground*, o viceversa. Esta corrección permite mantener un modelo de fondo más representativo para clasificar los píxeles del próximo *frame*.

4 Clasificación por texturas

La clasificación por texturas es un área importante de investigación en reconocimiento de patrones y visualización computacional. Entre ellos el descriptor LBP [12] es uno de los más usados debido a la simplicidad computacional y buena performance. El operador LBP original es invariante a los cambios en escala de grises y se calcula para un píxel determinado, tomando la diferencias de intensidades con los píxeles vecinos. Sin embargo, no es invariante a las rotaciones de imagen, por lo que distintas extensiones de este operador como LBROT, SILTP, entre otros incorporan la propiedad de invariancia en rotación [8][9]. En este trabajo se considera RSILTP [9] el cual se calcula en una región de 3x3 píxeles, tomando la diferencia de intensidad entre el centro del píxel p_c con respecto a sus 8 vecinos, según:

$$RSILTP(p_c) = \sum_{k=1}^{V} f(p_c, p_k) \tag{3}$$

$$f(p_c, p_k) = \begin{cases} 1, & \text{si } p_k > (1+\tau)p_c \\ 1, & \text{si } p_k < (1-\tau)p_c \\ 0, & \textit{en otro caso}, \end{cases} \tag{4}$$

donde p_c y p_k son los valores de intensidad del píxel central y el de su k-ésimo vecino, V es el número de vecinos, en este caso V=8, y el factor de escala τ que indica el rango de comparación. Para extraer la estructura fundamental de LBP, la idea de patrones uniformes se considera en el código binario al menos hasta dos transiciones de 0 a 1 o viceversa. Siguiendo esta convención, RSILTP sería menor igual a 2 en caso de texturas uniformes.

El indicador RSILTP (Ec. 3) permite identificar texturas con escala de intensidad similares, tanto para el frame de entrada, como para el modelo del fondo (muestras que ViBE selecciona aleatoriamente como representativas del fondo). En [13] se calcula el indicador RSILTP para el píxel de la imagen y el indicador RSILTP para el píxel correspondiente en el modelo de fondo, y se compara por la diferencia absoluta entre estos indicadores. De esta manera, puede suceder

que la imagen y el fondo tengan texturas uniformes pero tengan distinta escala de intensidad entre ellas. A diferencia, en este trabajo, p_c es el píxel a clasificar de la imagen y se lo compara en RSILTP con 8 muestras de vecinos p_k en el modelo í-esimo de fondo m_i. Esto hace que se compare el píxel de la imagen con la textura del fondo, y no de la misma imagen, con un solo indicador. Se utilizó valor de τ=0.2.

$$Reclasificar(p_c) = \begin{cases} 1, & \text{si } RSILTP(p_{c)} \leq 2 \textit{ en zona objeto detectado} \\ 0, & \textit{en otro caso}, \end{cases} \tag{5}$$

Las zonas de objetos detectados se refieren a zona clasificada como *foreground* según la Ec. 2. Esto se puede realizar por el procesamiento por separado de intensidad y luego por textura, ya que se descartan aquellas zonas que son clasificadas como *background* por ViBE y no son cercanas a un objeto en movimiento. Por ejemplo, se descarta procesar y reclasificar por textura fondo que es invariable en el tiempo.

En la Figura 3 se muestra un *frame* del video *Highway* y *Pedestrians* [15]. A la derecha en cada fila de la Figura 3 se muestra el resultado de la sustracción de fondo con ViBE propuesto con diferentes características, y las correcciones (a la derecha). Se puede observar en el caso de la persona caminando que no está conectada al cuerpo a la altura de la cintura (falsos negativos), y en el caso del vehículo de la derecha tiene también falsos negativos en la zona del vidriado. Los errores en la clasificación en el *frame* de *Highway* se producen en zonas donde la textura y color de la ruta es similar a la del auto, y también tiene problemas de clasificación en la zona donde se proyecta la sombra de los árboles a la izquierda de la imagen. Estos errores de falsos negativos fueron corregidos por el método propuesto. Sin embargo, la corrección sobre píxeles correspondientes a la sombras de los autos (falsos positivos) no son detectadas con la incorporación de textura en el modelo. Este problema también se puede notar en las imágenes de *Highway* mostradas para los métodos de [13] y [7].

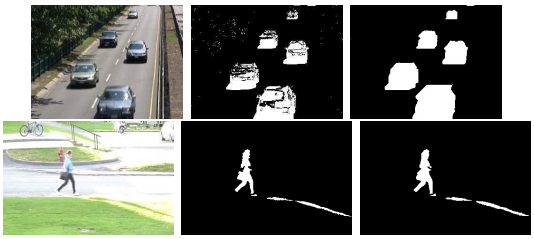

Fig. 3. Sustracción de fondo por multicaracterística. Imagen original (izq.), máscara ViBE por color (centro) y ViBE por color y textura (der). para *Highway* y *Pedestrians*

5 Resultados

Se evalua la performance del método propuesto sobre la base de datos *Change detection* (CDnet 2012) [15][16], la cual es usada por los métodos citados en el estado de arte. Se realizan los experimentos sobre una categoría de mayor interés de investigación en proyectos de tránsito vehicular y peatonal. En particular se seleccionaron los videos *Pedestrians*, *Highway* y *Office* de la categoría *Baseline*, *Park* de la categoría *Thermal*, *Overpass* de la categoría *Dynamic Background* y *Backdoor* de la categoría *Shadow* [15] [16], los cuales fueron adquiridos con cámaras estáticas para uso en el contexto de la videovigilancia. Estos videos cuentan con imágenes *groundtrue* para calcular la tasa de aciertos y comparar resultados. De esta forma, la librería de algoritmos de sustracción de fondo clásicos permite comparar con los métodos y contribuciones recientes en el área. En nuestro caso, se analizan los resultados obtenidos con otros métodos estocásticos tradicionales como ViBE+ y Gaussian Mixture Model (GMM)[5][13].

Para cada método se utilizaron tres métricas de evaluación conocidas *Precision*, *Recall* y *F-Measure* [15]. El indicador *F-Measure* se calcula con los indicadores *Precision* y *Recall*. El resultado es mejor, cuanto más cercana a 1 es la métrica. En la Tabla 1, se muestran los valores de estos indicadores para los videos de la categoría *Baseline*, y en la Tabla 2 se muestran los resultados para los videos de las restantes categorías .

Métodos /Videos	Highway			Pedestrians			Office		
	Re	Pr	FM	Re	Pr	FM	Re	Pr	FM
MoG [5]	0.92	0.93	0.92	0.99	0.92	0.95	0.49	0.75	0.59
MoG+texture [13]	0.93	0.80	0.86	0.97	0.73	0.84	0.99	0.54	0.70
ViBE+ [7]	0.93	0.93	0.93	0.95	0.96	0.96	0.70	0.92	0.80
Propuesto ViBE+texture	0.94	0.93	0.94	0.82	0.93	0.86	0.74	0.91	0.81

Tabla 1. Valores de indicadores para cada método con videos de *Baseline*

Métodos /Videos	Backdoor			Park			Overpass		
	Re	Pr	FM	Re	Pr	FM	Re	Pr	FM
MoG [5]	0.85	0.51	0.64	0.64	0.81	0.71	0.83	0.92	0.87
MoG+texture [13]	0.97	0.74	0.84	0.91	0.61	0.73	0.99	0.28	0.43
ViBE+ [7]	0.84	0.87	0.86	0.45	0.91	0.61	0.84	0.93	0.88
Propuesto ViBE+texture	0.85	0.95	0.90	0.49	0.93	0.64	0.87	0.84	0.86

Tabla 2. Valores de indicadores para cada método con otras categorías de la base *Change Detection*

Se puede observar para los videos *Highway*, *Office* y *Backdoor* que con el método propuesto ViBE+textura supera con el indicador FM a los otros

métodos. Sólo en el caso de *Park* el método ViBE con multicaracterística propuesto con el indicador FM no logra superar los valores de referencia de MoG [13], lo cual es muy satisfactorio. Es muy importante que los objetos detectados queden sin falsos negativos en su interior, porque esto hace que muchas veces se divida el objeto, causando dificultades en la etapa de seguimiento o *tracking* de objetos, ya que se consideran distintos *blobs* para un mismo objeto. En esta detección con modelo multicaracterística, la textura no ha podido separar correctamente la sombra de los objetos, sin embargo esto también se observa en los resultados de [13]. Es por eso que se propone en futuros trabajos incorporar módulos especializados para eliminar mejor la sombra de los objetos además de la textura, esperando que los valores obtenidos en las Tablas 1 y 2 con el método propuesto aumenten considerablemente.

6 Conclusiones

Se han mostrado los resultados preliminares del método de detección con un modelo de multicaracterísticas aplicado luego de la detección de objetos por color con el método ViBE los cuales han sido promisorios. Se han podido procesar los videos, y visualizar la clasificación en tiempo real, ya que el algoritmo tiene bajo costo computacional. Además, se ha logrado disminuir el error en la clasificación respecto de los métodos tradicionales, y se espera que con la separación de la sombra aumeten los valores de las métricas logradas. En futuros trabajos se pretende incorporar la adaptación automática de los umbrales como el valor de τ, teniendo en cuenta la variación de intensidad al procesar el video, ya que el algoritmo tiene inconvenientes en detectar texturas muy parecidas de sombra y del objeto en tonos oscuros, y comparar los resultados cuantitativamente con otros métodos utilizados actualmente. La actual versión está implementada en C# y en forma secuencial, y se prevé realizar una próxima versión utilizando GPU para aumentar la resolución de la imagen y reducir el tiempo computacional.

References

1. Kruegle H., CCTV Surveillance: Video practices and technology, Butterworth-Heinemann, (2014).
2. Gervasoni L., D'amato J., Barbuzza R., Vénere M.: Un método eficiente para la sustracción de fondo en videos usando GPU. In: Mecánica Computacional, Vol 33, pp. 1721–1731, ISSN 1666-6070, AMCA, Buenos Aires, (2014)
3. Dominguez L., Perez A., Rubiales, A., D'amato J., and Barbuzza R.: Herramientas para la detección y seguimiento de personas a partir de cámaras de seguridad. In: XXII Congreso Argentino de Ciencias de la Computación, pp. 251–260, (2016).
4. Barbuzza R., Dominguez, L.,Perez A., Esteberena L., Rubiales A., D'amato J., A Shadow Removal Approach for a Background Subtraction Algorithm, Computer Science CACIC 2017, 2018, Springer International Publishing,Cham, pp101-110
5. Legua C.,Seguimiento automático de objetos en sistemas con múltiples cámaras (2013).

6. Barnich O. and Van Droogenbroeck M.: ViBE: A powerful random technique to estimate the background in video sequences. In: 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, ISSN 1520-6149, pp. 945–948. (2009)
7. Barnich O. and Van Droogenbroeck M.: ViBE: A Universal Background Subtraction Algorithm for Video Sequences. In: IEEE Transactions on Image Processing, ISSN 1057-7149, 20 (6), pp. 1709–1724. (2011)
8. Heikkilä, M, Pietikainen and Heikkilä : A texture based-method for detection moving objects. In: 2010 Proc Brit Mach Vis. Conference vol(28), 4, pp. 21.1 (2004)
9. Heikkilä, M, Pietikäinen and Heikkilä : Modeling pixel process with scale invariant local patterns for background substraction in complex scenes. In: 2010 Proc IEEE Comp. Soc. Conf. Comput Vis. Pattern Recognition, Jun 2010, pp. 1301-1306 (2010)
10. Yeh C. Lin C Muchtar K, Kang L, :Real time background modeling based on multi-level texture description, In:Jun 2014 Inf Sci, vol (269) , pp. 106-127 (2014)
11. Wixson L., Detecting salient motion by accumulating directionality consistent flow, In:IEEE Trans. Pattern Anal Mach Intell, vol 22, 8, pp 774-780 (2000)
12. Ojala T, Pietikläinen M., Mäenpää T., Multiresolution gray-scale and rotation invariant texture classification with Local binary pattern, . In: IEEE Trans. Pattern Anal Mach Intell, vol 24, 7, pp 971-987 (2002)
13. D. Yang, C. Zhao, X. Zhang and S. Huang,; Background Modeling by Stability of Adaptive Features in Complex Scenes, In:IEEE Transactions on Image Processing, vol. 27, no. 3, pp. 1112-1125, March 2018.doi: 10.1109/TIP.2017.2768828
14. K. Roy, M. R. Arefin, F. Makhmudkhujaev, O. Chae and J. Kim;Background Subtraction Using Dominant Directional Pattern, IEEE Access, vol. 6, pp. 39917-39926, 2018.doi: 10.1109/ACCESS.2018.2846749
15. Goyette N., Jodoin P. M.,Porikli F.,Konrad J. and Ishwar P.: Changedetection.net: A new change detection benchmark dataset. In: 2012 IEEE Comp. Soc. Conference on Computer Vision and Pattern Recognition Workshops, 20 (6), pp. 1–8, (2012)
16. Sobral A. Vacavant A. A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos. In: Computer Vision and Image Understanding, Vol. 122, Elsevier, pp. 4–21, (2014)