

Separación de sombras a los objetos detectados con sustracción de fondo en video

Rosana Barbuzza^{1,3}, Leonardo Fernandez Esteberena^{1,2}, Leonardo Dominguez^{1,2}, Alejandro Perez¹, Aldo Rubiales^{1,3} y Juan P. D'Amato^{1,2}

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. El trabajo presenta los resultados preliminares de un algoritmo de detección y eliminación de sombras, en secuencias de video. Se propone que a partir de la base de la sustracción de fondo con el algoritmo *Visual Background Extraction (ViBE)*, que identifica zonas de movimiento, aplicar un post-procesamiento para separar los píxeles del objeto real y los de sombra. Debido a que las zonas de sombras tienen características parecidas a la de los objetos en movimiento, la separación es una tarea difícil, por lo que los algoritmos usados para esta clasificación producen gran cantidad de falsos positivos. Para atacar este problema, se parte de la premisa de utilizar información del objeto, como el tamaño y la dirección de movimiento, para estimar la posición aproximada de la sombra. Además, se realiza el análisis de las similitudes entre el cuadro actual y el modelo de fondo, con el indicador tradicional de correlación cruzada normalizada para detectar sombras. El algoritmo se puede aplicar para detectar personas o vehículos en aplicaciones para seguridad ciudadana, monitoreo de tráfico, análisis deportivo, entre otros. Los resultados obtenidos en la detección de objetos muestran que es factible recortar la sombra con alta tasa de acierto y con bajo costo computacional, lo cual también permite mejorar etapas de procesamiento posteriores, como el reconocimiento y el seguimiento de los objetos.

Palabras clave: procesamiento de video, detección de objetos, segmentación

1 Introducción

Actualmente, existe un gran interés tanto en la industria como en la investigación en contar con sistemas de video análisis. En el área de seguridad, se ha mostrado que el video digital es una herramienta fundamental tanto para la pericia forense, como para la disuasión del delito. En particular, en Argentina la inseguridad es un tema crítico que requiere su atención inmediata, para lo cual ha crecido la cantidad de centros de monitoreo, con gran cantidad de operadores observando cámaras. Ciertamente, este mecanismo no es eficiente por lo que es importante contar con técnicas de análisis de video que ayuden a los operadores en las diferentes tareas cotidianas. En esta línea, este grupo de investigación

ha presentado varios trabajos, principalmente mostrando la arquitectura de un sistema abierto, distribuido y escalable [1].

La complejidad de estos algoritmos radica en trabajar en ambientes dinámicos, como son las cámaras del exterior, donde sufren las inclemencias del clima. Para contrarrestar estos efectos, se suelen utilizar ventanas de movimiento como se menciona en [2], cuyo objetivo consiste en aplicar una operación lógica, para procesar únicamente los movimientos que se detecten dentro de sus dimensiones y descartar todos aquellos que se produzcan por fuera.

Estos algoritmos de seguimiento finalmente son utilizados en plataformas de video-análisis en virtud de detectar y analizar situaciones de interés, por lo que es crítico tener una alta tasa de confianza en la detección. En este contexto, la sombra afecta considerablemente la percepción de los objetos detectados, tal como se ve en la Figura 1, ya que altera el tamaño, color u orientación del mismo afectando drásticamente una posterior clasificación por tamaño o color; por lo que debe reducirse de alguna manera.



Fig. 1. Imágenes de diferentes casos, donde la sombra cambia el tamaño y forma del objeto.

El presente trabajo se enfoca en poder detectar y eliminar sombras a partir de videos en escala de grises o en el espacio color transformado a escala de gris (ej. `rgb2gray` o HSV correspondiente a la componente Value). Esta consideración también es adecuada ya que muchas cámaras de bajo costo a menudo no proveen una imagen de calidad suficiente, afectada también por la compresión de video. Lo novedoso del método propuesto es que utiliza información del objeto y de su orientación, para determinar *a priori* la posición de la sombra, para luego aplicar en la región externa al objeto un método tradicional que tiene en cuenta las características de textura y color en las zonas de sombras.

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 se realiza la propuesta de detección de sombras y de resultados obtenidos. Finalmente en la sección 6 se presentan las conclusiones y futuros trabajos.

2 Estado del arte

Las técnicas de detección de eventos usados en sistemas de *Video-Surveillance*, se basan principalmente en discriminar rápidamente el movimiento a partir de una video-cámara fija. Trabajos como [3], detallan los algoritmos más comunes para la detección y seguimiento de objetos. En [4] se realiza una comparación interesante entre distintos algoritmos de detección básicos, en donde se concluye que la combinación de éstos puede ser muy útil para disminuir la tasa de falsos positivos, mientras se mantiene el tiempo y la tasa de verdaderos positivos.

Los sustractores de fondo usualmente usados clasifican las sombras como parte de los objetos, lo cual altera el tamaño y forma de los mismos, afectando en consecuencia la eficacia de tales algoritmos [5]. También, dificulta otros procesamiento que necesitan el resultado de la detección de objetos, como por ejemplo, la clasificación, o el seguimiento de la trayectoria y análisis de comportamiento (ej. merodeo, vandalismo, infracciones de tránsito, etc.). Este problema también afecta a técnicas basadas en características [6].

Existen varias investigaciones acerca de la detección y separación de sombras. Algunos algoritmos que trabajan sobre imágenes estáticas, son computacionalmente complejos y no aplicables para análisis de video en tiempo real [7]. A diferencia, otros algoritmos están diseñados especialmente para video en escala de grises o color, y tienen menor costo computacional. Para realizar la clasificación, la mayoría se basa en las características usuales de la región con sombras (zonas más oscuras que el fondo de la escena, color o textura uniforme e invariable, etc.). Aunque estas características permiten crear candidatos para determinar la zona de sombras, no son decisivas, lo cual causa que todos los métodos fallen en la clasificación, ya que estas propiedades suelen aparecer también sobre partes de los objetos [8][9]. En particular, otro grupo de algoritmos incluyen información sobre la geometría de la sombra o modelo de iluminación, algunos especializados solamente en sombra de personas [10].

3 Sustractor de fondo en video

En este trabajo se utiliza el sustractor de fondo ViBE, propuesto en [11], el cual tiene buen comportamiento en entornos variados como suelen ser típicos en videovigilancia. Entre las virtudes de este método 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 [3]. En la Figura 2, se muestra la sustracción de fondo realizada con ViBE para un *frame* particular de video de monitoreo urbano. En el ejemplo, se puede ver la imagen original y a su izquierda, las componentes *foreground* detectadas, personas y vehículos. La proyección de la sombra del ómnibus forma parte del objeto, deformando notablemente la silueta del mismo, y el tamaño se incrementa en aproximadamente un 50%.

El algoritmo ViBE utiliza un modelo representativo del fondo basado en píxeles (guarda aleatoriamente muestras del valor de intensidad de cada píxel

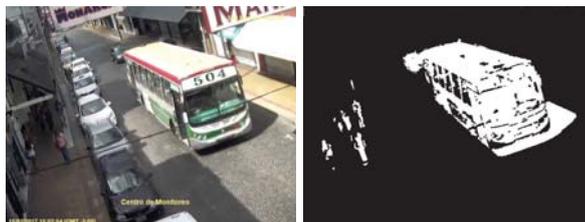


Fig. 2. Imagen del video (*izq.*) y sustracción del fondo con ViBE (*der.*), para un video registrado en cámara de monitoreo.

de la secuencia previa de video). En la Figura 3, se representa la secuencia de pasos del algoritmo ViBE original y los módulos incorporados para mejorar la clasificación. El módulo *Detección*, clasifica las escenas como *background* o *foreground*, calculando la distancia entre cada píxel de una imagen actual respecto de las muestras guardadas en 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, por un nuevo valor de intensidad del mismo píxel en el *frame* actual (Figura 3 módulo *Actualizar modelo*). La aleatoriedad sobre este mecanismo de actualización permite reducir gradualmente con el tiempo, la probabilidad de que las muestras guardadas persistan en el modelo. Es muy importante que solamente aquellos píxeles procesados y clasificados como *background* reemplacen muestras en el modelo de fondo correspondiente. La inserción de píxeles mal clasificados, que pertenecen a objetos en movimiento (*foreground*), o que resultan inciertos en la clasificación, pueden alterar significativamente los resultados de la detección. Esto último es muy importante ya que ha permitido la introducción de mejoras al método ViBE original logrando mayor tasa de detección, por ejemplo para adaptar el algoritmo a los movimientos de escena o fondo dinámico (detectados como falsos positivos), o el rellenado de objetos detectados en forma incompleta (falsos negativos) [12][13]. En este trabajo, se agrega el algoritmo de operación morfológica de cierre de los objetos (dilatación, rellenado de agujeros), para lograr tener componentes conectadas, se seleccionan los objetos con tamaño considerado, y se descartan los pequeños y aislados (Módulo *Contornos y Rellenado*). Luego de este proceso, se aplica el algoritmo de separación de sombras del objeto real (Módulo *Sombras*).

Este módulo propone corregir los píxeles clasificados como *foreground* pero que se corresponden a sombras. Cabe aclarar también, que no resulta adecuado clasificar las sombras como *background*, ya que estas muestras no deberían modificar o alterar el modelo de fondo, cuando actualiza el modelo, para evitar errores posteriores en la clasificación.

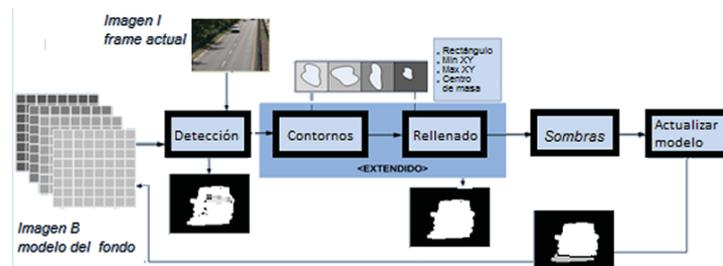


Fig. 3. Algoritmo ViBE y módulos de mejoras implementados

4 Propuesta de detección de sombras

Para la detección y eliminación de sombras por un lado se utiliza la correlación cruzada normalizada (NCC) entre la imagen actual (I) y una imagen representativa del fondo (B), ajustando con propiedades características de la sombra como en [14], pero además se incorpora al método original de NCC, conocimiento elemental sobre los objetos para mejorar la tasa de detección, detallando ambos a continuación.

4.1 Identificación de píxeles candidatos con NCC

El indicador NCC permite identificar imágenes similares con diferente escala de intensidad. Para representar al fondo (B), se utilizó una imagen con las muestras que ViBE selecciona aleatoriamente como representativas del modelo del fondo (Ver Figura 3).

Para cada píxel (i, j) clasificado como *foreground* por ViBE, se consideran los vecinos en una región R cuadrada, con centro en ese píxel. Los píxeles candidatos a clasificar como sombra son aquellos que tienen valor de correlación alto, teniendo en cuenta esta región de vecinos:

$$C(i, j) = \sum_{i=-N}^N \sum_{j=-N}^N (I(i, j) * B(i, j)) \quad (1)$$

donde $I(i, j)$ y $B(i, j)$ corresponde al valor de intensidad del píxel en la imagen I y B, respectivamente, y $(2N+1)^2$ es el tamaño de R. La correlación se normaliza según:

$$NCC(i, j) = \frac{C(i, j)}{\sqrt{mI(i, j)} * \sqrt{mB(i, j)}} \quad (2)$$

donde $mI(i, j)$ y $mB(i, j)$ son los momentos no centrados de orden 2 en la misma región R centrada en el píxel (i, j) , en la imagen I y B, respectivamente.

El valor de referencia para detectar sombras con NCC (Ec. 2) es un valor alto definido en el rango [0.95-0.98] [14], y se usa como paso inicial para detectar posibles candidatos.

Luego, se rectifican píxeles, usando estadística sobre la región R teniendo en cuenta el análisis de la relación entre los valores de intensidad de las imágenes I y B para cada pixel (i, j) :

$$\alpha(i, j) = \frac{I(i, j)}{B(i, j)} \quad (3)$$

Las zonas de sombras deben ajustarse a un rango definido de valores, donde α en la Ec. 3 se define usualmente entre $[0.4-1]$, dada la oscuridad de la sombra sobre el modelo de fondo [14]. También se considera, que la desviación estándar de la relación α calculada en la región R de vecinos, debe tener un valor bajo debido a que la zona de sombra es un área homogénea. Se recomienda un desvío estándar menor a 0.05 en la región R de tamaño 5×5 píxeles, siendo N igual a 2. El problema del método original [14] es que genera falsos negativos cuando el valor de intensidad del objeto coincide con los valores de la sombra.

En la Figura 4 se muestra un *frame* del video Pedestrians [15] (izquierda) y otro *frame* de un video de monitoreo urbano (derecha). En la primer fila de la Figura 4 se muestra el resultado de la sustracción de fondo con ViBE. Se puede observar en el caso de la persona caminando, que la sombra (falsos positivos) no está conectada al cuerpo y en el caso del vehículo la sombra se proyecta en la parte inferior, desformando el objeto. En la segunda fila de la Figura 4, se utilizaron parámetros recomendados para NCC (Ec. 2) mayor a 0.95, y para α en el rango $[0.4-1]$. Se marca la zona de sombras resultantes (en color amarillo). Se puede observar que no siempre la detección de sombra es correcta, como por ejemplo, sobre las piernas de la persona, y en los vidrios del vehículo, generando falsos negativos. El método no trabaja correctamente, principalmente en las escenas donde hay reflejos de luz, o el fondo se parece mucho a la sombra. En estos casos, fue necesario considerar diferentes valores de α y desvío estándar, dependiendo del ruido en la señal de video, y en casos donde la sombra es muy oscura.

4.2 Incorporación de conocimiento

En este trabajo se tuvo en cuenta no aplicar las reglas anteriores para NCC en zonas potenciales del objeto verdadero. En primer lugar, la clasificación de sombras se realiza a nivel de una región rectangular o *blob* detectado (componentes conectadas en la máscara *foreground* detectada por ViBE) y no de cada píxel aislado como en la versión original del método [14]. Se propone entonces, utilizar la información espacial dentro de esta región en la imagen I . En el caso de trabajar con videos de personas y/o vehículos a partir de video con cámara fija, se decidió además incorporar conocimiento de algunas propiedades de los mismos (ej. orientación, forma, tamaño).

En particular, al encerrar a las personas o los vehículos con una elipse, usando el centro de masa de la máscara *foreground*, es muy probable que la proyección de la sombra exceda los límites de la elipse, como en el caso del vehículo en la Figura 4. En otros casos, la proyección de la sombra puede quedar separada en una elipse diferente como en el caso de la persona caminando.



Fig. 4. Sustracción de fondo y detección de sombra. Imagen original (izq.) y Máscara resultante (derecha). Algoritmo ViBE (arriba). Algoritmo detección de sombras con NCC (centro) y Detección de sombras con algoritmo propuesto (abajo)

En cada *blob*, se calcula la distribución direccional para encontrar la elipse que mejor se ajusta a la distribución de píxeles *foreground*. Como resultado de este método, se define una elipse mediante el eje mayor, el eje menor, y la orientación del eje mayor respecto del eje horizontal [16]. Además, fue necesario adaptar estos valores para que cubran la totalidad de los diferentes objetos (tanto personas como vehículos), y no exceder demasiado los límites del *blob*. En forma heurística, resultó adecuado un factor de 1.8 para el eje mayor y de 1.6 para el eje menor.

Como primer intento de no invadir en el recorrido las zonas del objeto verdadero, se tuvo en cuenta que la altura y el ancho del cuerpo humano mantienen cierta proporción, y además que la persona camina erguida. De esta manera, la regla de decisión para determinar que en un *blob* hay una persona, es por el tamaño del área, la relación entre el diámetro mayor y menor de la elipse sea mayor igual a 3, y finalmente que la orientación del eje mayor se mantenga cercana a los 90. En el caso de la Figura 4 (arriba), se puede ver que la diagonal mayor de la elipse es casi vertical para el caso de la persona, y que la diagonal mayor de la elipse que encierra la sombra queda casi horizontal. Para identificar automóviles, se consideró que la orientación puede variar, sin embargo el tamaño del *blob* es mucho mayor que el que contiene personas, bicicletas o motos. En la Figura 4 (abajo), se puede ver que los resultados del algoritmo propuesto, mejoran respecto de algoritmo original de ViBE+ (Figura 4, arriba) y también de ViBE+[14] (Figura 4, centro). La persona que camina no tiene sombras sobre las piernas, y tampoco se observan sombras en los vidrios del vehículo (falsos negativos). Cuando se detecta sombra fuera de la elipse, el algoritmo continúa

sobre los píxeles vecinos, manteniendo la continuidad incluso si sobrepasa los límites de la elipse.

5 Resultados

El algoritmo propuesto fue testeado con los videos reales *Pedestrians* y *Highway* de la base de datos [15] [17], 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)[3].

Para cada método se utilizaron dos métricas evaluación conocidas *Precision*, *Recall* según:

$$Precision = \frac{TP}{(TP+FP)} \quad (4)$$

$$Recall = \frac{TP}{(TP+FN)} \quad (5)$$

siendo FN el número de falsos negativos, FP el número de falsos positivos, TP es el total de positivos y TN el total de negativos. En este caso, el resultado es mejor, cuanto más cercana a 1 es la métrica. También se consideró un indicador sobre errores en la sombra como,

$$TShadow = \frac{nbErrorShadows}{FP} \quad (6)$$

siendo nbErrorShadows el total de Falsos Positivos que se producen en zona de sombras según el *groundtrue*, y FP el número de falsos positivos.

En la Tabla 1, se muestran los valores de las ecuaciones anteriores para aproximadamente 700 imágenes del video *Pedestrians*, y en la Tabla 2 los valores para 1300 imágenes de *Highway*.

Métodos	Recall	Precision	Tshadow
GMM [18]	0.98	0.93	0.22
ViBE+ [11]	0.93	0.96	0.50
ViBE+NCC [14]	0.83	0.99	0.11
Propuesto ViBE+NCC+Información	0.91	0.99	0.07

Table 1. Valores de indicadores para cada método con el video Pedestrians

Se puede observar para ambos videos *Pedestrians* y *Highway*, que con los método ViBE+ o GMM, el porcentaje TShadow es mucho mayor (cercano al

Métodos	Recall	Precision	Tshadow
GMM [18]	0.89	0.91	0.81
ViBE+ [11]	0.84	0.92	0.95
ViBE+NCC [14]	0.58	0.94	0.47
Propuesto ViBE+NCC+Información	0.85	0.94	0.55

Table 2. Valores de indicadores para cada método con el video Highway

doble) que con el método propuesto. El método ViBE+NCC[14] aumenta la cantidad de falsos negativos haciendo que la métrica *Recall* disminuya considerablemente respecto de ViBE+ (en *Pedestrians* de 0.93 a 0.83 y en *Highway* de 0.84 a 0.58). También en ambos videos con el método propuesto, se nota un aumento de la tasa *Precision* respecto de ViBE+, indicado una reducción de la cantidad de falsos positivos. Finalmente, con el método propuesto ViBE+NCC+información, la tasa *Recall* y *Precision* es comparable a la de los métodos ViBE+ y GMM, sin embargo, el porcentaje de mal clasificados en zona de sombras con *TShadows* es notablemente menor.

6 Conclusiones

En este trabajo, se han mostrado los resultados preliminares del método de detección y separación de la sombra, basado en información contextual, los cuales han sido promisorios. Se ha podido procesar los videos, y visualizar la clasificación en tiempo real, ya que algoritmo tiene bajo costo computacional.

Se ha logrado disminuir el error en la clasificación respecto de la aplicación de NCC tradicional, calculando además la orientación y ubicación probable del objeto, la determinación de persona o vehículo.

En futuros trabajos se pretende incorporar la adaptación automática de los umbrales del rango para α , teniendo en cuenta la variación de intensidad de la sombra al procesar el video, ya que el algoritmo tiene inconvenientes en detectar zonas de sombra muy oscuras. El otro desafío es que los parámetros del algoritmo se adapten automáticamente a diferentes momentos del día y al mismo tiempo, comparando cuantitativamente con otros métodos de referencia.

References

1. Dominguez L., Perez A., Rubiales, A., Damato J., and Barbuzza R.: Herramientas para la detección y seguimiento de personas a partir de cámaras de seguridad. In: Proceedings del XXII Congreso Argentino de Ciencias de la Computación, pp. 251–260, (2016).
2. Kruegle H. ,CCTV Surveillance: Video practices and technology, Butterworth-Heinemann, (2014).
3. Legua C. C.,Seguimiento automático de objetos en sistemas con múltiples cámaras (2013).
4. Shaikh S. H. and Saeed K. and Chaki N. ,Moving Object Detection Using Background Subtraction, Springer (2014).

5. Azab, M., Shedeed, H., Hussein, A., A new technique for background modeling and subtraction for motion detection in real-time videos. In: IEEE International Conference on Image Processing (ICIP), pp. 3453-3456 (2010).
6. Raad Ahmed Hadi, Ghazali Sulong, and Loay Edwar George; Vehicle detection and tracking techniques: A concise review, Signal and Image Processing : An International Journal (SIPIJ) Vol.5, No.1, (2014).
7. Fredembach, C., Finlayson, G.,: Simple shadow removal. In: Proceedings of the 18th International Conference on Pattern Recognition (ICPR06), pp. 832–835, (2006)
8. Qin R. and S. Liao and Z. Lei and S. Z. Li: Moving Cast Shadow Removal Based on Local Descriptors. In: 2010 20th International Conference on Pattern Recognition, pp. 1377–1380, (2010)
9. Chen F., Zhu and B., Jing W. and Yuan L. : Removal shadow with background subtraction model ViBE algorithm. In: 2013 2nd International Symposium on Instrumentation and Measurement, Sensor Network and Automation (IMSNA), pp. 264–269,(2013)
10. Bi S., Liang D., Shen X. and Wang Q.: Human Cast Shadow Elimination Method Based on Orientation Information Measures. In: IEEE International Conference on Automation and Logistics, Jinan, pp. 1567-1571, (2007)
11. 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)
12. Gervasoni L., Damato 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)
13. Barbuzza R., Damato J., Rubiales A. Dominguez L, Perez A., and Venere M.: Un método para la sustracción de fondo en videos inestables. In: Mecánica Computacional, Vol 34, pp 3409–3417, (2016)
14. Jacques J. C. S., Jung C. R. and Musse S. R.: Background Subtraction and Shadow Detection in Grayscale Video Sequences. In: XVIII Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI'05), pp. 189–196. IEEE Computer Society, ISSN 1530-1834, Washington DC, USA (2005)
15. Goyette N., Jodoin P. M., Porikli F., Konrad J. and Ishwar P.: Changedetection.net: A new change detection benchmark dataset. In: 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 20 (6), pp. 1–8, (2012)
16. Lahoz Beltra, R., Ortega J. and Fernandez Montraveta C.: Métodos estadísticos en biología del comportamiento: Editorial Complutense, ISBN 84-7491-512-0, (1994)
17. 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)
18. Sobral A. Vacavant A. Improved adaptive Gaussian mixture model for background subtraction. In: Proc. Int. Conf. Pattern Recognition, IEEE, Piscataway, NJ, pp. 28–31 (2004)