

Capítulo 3

Enfoque propuesto para la detección del humo de señales de video.

3.1 Comportamiento del enfoque propuesto.

Una visión general del método propuesto se muestra en la figura 2. El método genera una entrada subsecuente de video conocido como fragmento de código mediante un muestreo uniforme de la secuencia del video original.

En cada cuadro de la muestra hay una partición en tres regiones diferentes (humo, cielo y tierra) que se realiza con un proceso de clasificación parche por parche. Cada parche de imagen está clasificado por la extracción de un vector de características del parche, y el proceso del vector de características asigna una determinada clase al parche.

El proceso de segmentación de imágenes también puede ser ejecutado pixel a pixel al procesar el parche alrededor del pixel de interés, pero a expensas de un aumento del tiempo computacional.

Después de la segmentación de un frame, una reclasificación de un parche se lleva a cabo teniendo en cuenta la interacción espacial y temporal entre los parches vecinos (cuboide espacio-temporal).

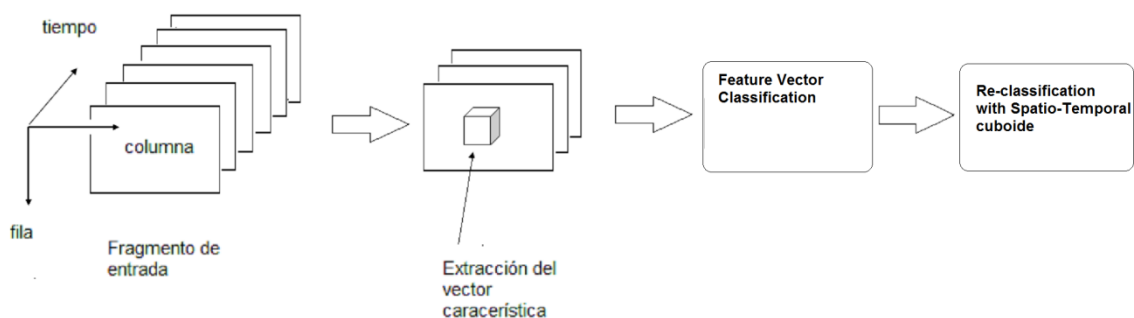


Figura 2 Segmentación de fragmentos para la detección de incendios forestales

3.2 Extracción de características.

La estructura del diccionario es flexible en el sentido de que es fácil de considerar diferentes tipos de características. Los valores de intensidad de color en un parche se utilizan como características por concatenación de los planos RGB del parche en un único vector multiplicando por 3 la dimensión del parche.

El uso de la información de color como característica es fundamental para discriminar entre el humo, el cielo y las regiones terrestres. A pesar de que un vector de rasgos se extrae de un parche de la imagen, parte de la información proviene de otros parches con adyacencia temporal para dar cuenta de la información temporal.

La incursión de la información temporal se basa en el hecho de que los cambios en regiones de humo y la firmeza de las regiones del cielo y de tierra se incrustan en las características temporales.

Un vector de rasgos puede ser visto como la concatenación de un cuboide espacio-temporal de píxeles de color en un único vector unidimensional. La información estadística también se incluye como características mediante el cálculo de la media y la varianza de cada plano de color en un parche.

Un vector característica $x_i(s_i, t_j) \in R^N$ es asociado a un parche de imagen $p_i(s_i, t_j)$ en una locación espacial s_i y una locación temporal t_j . Las N entradas de un vector de características consiste de la intensidad del valor del color sobre los 3 planos RGB del parche $p_i(s_i, t_j)$ que son aplicados como un vector columna.

$$y(s_i, t_j) = \left[y_R(s_i, t_j)^T, y_G(s_i, t_j)^T, y_B(s_i, t_j)^T \right]^T \in R^{3N} \quad (20)$$

Valores de color de intensidad de otros parches con adyacencia temporal localizados en fotogramas en las instancias de tiempo t_{j-1} y t_{j+1}

$$\begin{aligned} & \left[\mathbf{y}(\mathbf{s}_i, t_{j-1})^T, \mathbf{y}(\mathbf{s}_i, t_{j+1})^T \right]^T = \\ & \left[\mathbf{y}_R(\mathbf{s}_i, t_{j-1})^T, \mathbf{y}_G(\mathbf{s}_i, t_{j-1})^T, \mathbf{y}_B(\mathbf{s}_i, t_{j-1})^T, \mathbf{y}_R(\mathbf{s}_i, t_{j+1})^T, \mathbf{y}_G(\mathbf{s}_i, t_{j+1})^T, \mathbf{y}_B(\mathbf{s}_i, t_{j+1})^T \right]^T \in \\ & R^{6N} \end{aligned} \quad (21)$$

La media de los valores de la intensidad en cada plano de color RGB en la localidad \mathbf{s}_i y en diferentes estancias de tiempo t_{j-1}, t_j, t_{j+1} .

$$\begin{aligned} \boldsymbol{\mu} = & \left[\overline{\mathbf{y}_R(\mathbf{s}_i, t_{j-1})}, \overline{\mathbf{y}_G(\mathbf{s}_i, t_{j-1})}, \overline{\mathbf{y}_B(\mathbf{s}_i, t_{j-1})}, \overline{\mathbf{y}_R(\mathbf{s}_i, t_j)}, \overline{\mathbf{y}_G(\mathbf{s}_i, t_j)}, \overline{\mathbf{y}_B(\mathbf{s}_i, t_j)}, \overline{\mathbf{y}_R(\mathbf{s}_i, t_{j+1})}, \right. \\ & \left. \overline{\mathbf{y}_G(\mathbf{s}_i, t_{j+1})}, \overline{\mathbf{y}_B(\mathbf{s}_i, t_{j+1})} \right]^T \in R^9 \end{aligned} \quad (22)$$

Y las varianzas de las intensidades de cada plano de color en una locación espacial \mathbf{s}_i y en diferentes instancias de tiempo t_{j-1}, t_j, t_{j+1}

$$\begin{aligned} \boldsymbol{\sigma} = & \left[\text{var}(\mathbf{y}_R(\mathbf{s}_i, t_{j-1})), \text{var}(\mathbf{y}_G(\mathbf{s}_i, t_{j-1})), \text{var}(\mathbf{y}_B(\mathbf{s}_i, t_{j-1})), \text{var}(\mathbf{y}_R(\mathbf{s}_i, t_j)), \right. \\ & \text{var}(\mathbf{y}_G(\mathbf{s}_i, t_j)), \text{var}(\mathbf{y}_B(\mathbf{s}_i, t_j)), \text{var}(\mathbf{y}_R(\mathbf{s}_i, t_{j+1})), \text{var}(\mathbf{y}_G(\mathbf{s}_i, t_{j+1})), \\ & \left. \text{var}(\mathbf{y}_B(\mathbf{s}_i, t_{j+1})) \right]^T \in R^9 \end{aligned} \quad (23)$$

Teniendo en cuenta toda esta información, un vector característica es construido,

$$\mathbf{x}_i(\mathbf{s}_i, t_j) = \left[\mathbf{y}(\mathbf{s}_i, t_{j-1})^T, \mathbf{y}(\mathbf{s}_i, t_j)^T, \mathbf{y}(\mathbf{s}_i, t_{j+1})^T, \boldsymbol{\mu}, \boldsymbol{\sigma} \right]^T \in R^N \quad (24)$$

El número de entradas de cada vector de rasgos esta dado por $N = 3 [3 (\sqrt{n} \times \sqrt{n} + 2)]$. Los vectores características incorporan el color y la información temporal llegando a ser de una alta dimensión siendo este el motivo la elección de un parche de tamaño pequeño $n=3 \times 3$.

3.3 Clasificación

El clasificador basado en diccionarios de aprendizaje es una conexión en cascada de dos sistemas. La primera etapa es una transformación no lineal de la señal, $\mathcal{T}: \mathbf{R}^N \rightarrow \mathbf{R}^p$ con la prueba del vector característica, $\mathbf{x} \in \mathbf{R}^N$ como entrada y residuo del vector característica y $\mathbf{r} = [r_1(\mathbf{x}, \mathbf{D}_1), r_2(\mathbf{x}, \mathbf{D}_2), \dots, r_p(\mathbf{x}, \mathbf{D}_p)]^T \in \mathbf{R}^p$ como salida.

Dado que todas las entradas (características) en \mathbf{r} no son negativas, el resultado de esta etapa es un vector característica transformado, donde \mathbf{r} esta en el primer hiper-cuadrante en el espacio p-dimensional.

La segunda etapa consiste en un clasificador con \mathbf{r} como la entrada del vector característica con valores de pesos sinápticos en el conjunto $\{-1,1\}$, y con la función escalón unitario $u(v)$ como la función de activación. El hiper-plano con el i -ésimo clasificador se encuentra en el primer hiper-cuadrante en el espacio p-dimensional y viene dado por:

$$r_i(\mathbf{x}, \mathbf{D}_i) - \sum_{\substack{j=1 \\ j \neq i}}^p r_j(\mathbf{x}, \mathbf{D}_j) = 0. \quad (25)$$

Las características residuales a la salida de la etapa de transformación son encontradas para resolver el problema de la descomposición esparsiva en la ecuación (10).

Existen diferentes métodos en la literatura de la codificación esparsiva tales como los métodos de ajuste activo, los métodos en umbralización suave y homotapía.

3.4 Interacción espacial y temporal para la reducción de los parches mal clasificados.

Las entradas de un vector de rasgos transformado $\{r_i(\mathbf{x}, \mathbf{D}_i)\}_{i=1}^p$ son los errores de la representación esparsiva del correspondiente vector característica \mathbf{x} sobre diccionarios diferentes $\{\mathbf{D}_i\}_{i=1}^p$. Después de que las características transformadas son usadas para la clasificación de parches durante la detección de humo en una

secuencia de fragmentos un problema que surge es la mal clasificación de parches.

Con el fin de mejorar aun más la segmentación, un proceso de re-clasificación de parches se lleva a cabo por contabilidad para la interacción espacial y temporal entre los parches vecinos de la misma manera que el filtrado pasa-bajas utiliza múltiples píxeles vecinos para calcular un nuevo valor de pixel, lo que permite la reducción de los pixeles mal clasificados.

La reasignación de etiquetas de un parche clasificado se realiza ahora mediante la búsqueda de sus parches vecinos en los ámbitos espaciales y temporales (cuboide espacio-temporal de los parches) como se muestra en la figura 3, y la nueva clase que se asignará es el que tiene la más alta frecuencia de aparición dentro del cuboide espacio- temporal de los parches.

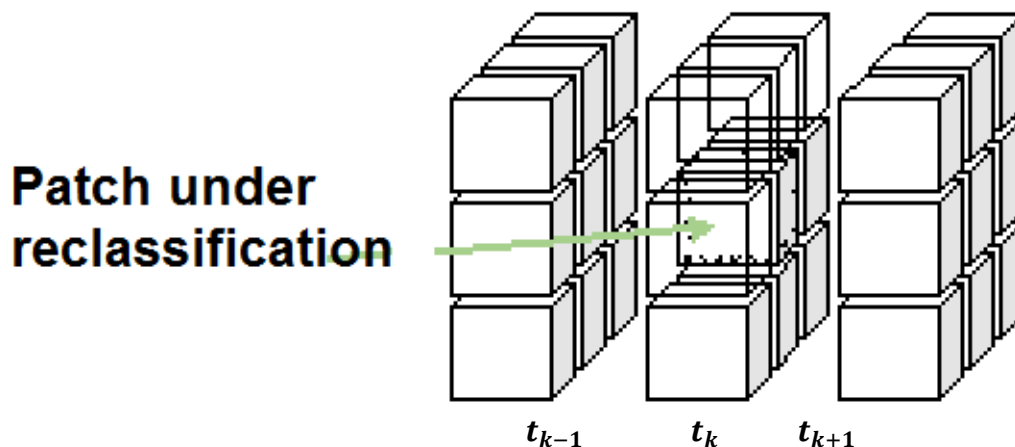


Figura 3 Estructura de parches vecinos a través de los dominios espaciales y temporales.

Para volver a clasificar un parche de interés $p(i, j, t_k)$ un cuboide espacio-temporal de parches es construido $P = \{p(i \pm q, j \pm q, t_{k \pm q}); q = 0, 1, \dots, Q\}$ con adyacencia entre parches vecinos sobre el espacio y tiempo.

El problema consiste en asignar una nueva etiqueta l_m del conjunto de etiquetas $L = \{l_1, l_2, \dots, l_p\}$ a un parche ya etiquetado $p(i, j, t_k)$. Se deja P_{l_m} que es un subconjunto de \mathbf{P} con elementos de parche que tienen la misma etiqueta l_m y con

$n(P_{lm})$ siendo la cardinalidad del subconjunto P_{lm} que es el número de elementos de parche que P_{lm} contiene.

La nueva etiqueta para $p(i, j, t_k)$ es asignada de acuerdo a:

$$l = \arg \max_{l_m} n(\mathcal{P}_{l_m}); m = 1, 2, \dots, p. \quad (26)$$

3.5 Diccionarios de aprendizaje.

El entrenamiento del clasificador basado en diccionarios de aprendizaje consiste en encontrar el conjunto de diccionarios $\{\mathbf{D}_i\}_{i=1}^P$ que es el conjunto de parámetros en la etapa de transformación del clasificador.

Hay 3 diccionarios ($P=3$) para ser entrenados por la representación esparsiva de los vectores característica extraídos del humo, cielo y regiones de tierra. Por cada cámara de vigilancia un diccionario es inicializado con vectores característica extraídos de una región de interés.

En lugar de aprender y almacenar grandes diccionarios de representación esparsiva de vectores característica extraídos de múltiples escenarios (para cada escenario hay una cámara de vigilancia) se utilizarán pequeños diccionarios que aprenden de forma independiente para cada cámara.

Los conjuntos de formación para el aprendizaje de diccionario se generan con las muestras tomadas de escenas del paisaje de interés capturado por una cámara de vigilancia de la misma manera que se inicializan los diccionarios.

Dado tres conjuntos de formación de las observaciones de humo, el cielo y el terreno y la versión inicial de los diccionarios \mathbf{D}_1 , \mathbf{D}_2 y \mathbf{D}_3 estos diccionarios están entrenados utilizando métodos K-SVD que es una extensión del algoritmo K-means para el diccionario de aprendizaje [10].

El método K-SVD fue elegido ya que ofrece ventajas sobre el método MOD [11]. K-SVD y MOD (Método de Direcciones Óptimas) puede trabajar con cualquier técnica de búsqueda, sin embargo, MOD requiere una matriz de inversión que lo hace más práctico cuando el número de átomos es muy grande.