APRENDIZAJE DE REGLAS DIFUSAS PARA LA CLASIFICACIÓN DE COMPORTAMIENTOS EN UN SISTEMA DE VIDEOVIGILANCIA COGNITIVA

Javier Albusac[1] **J.J Castro-Schez**[1] **David Vallejo**[1] **Luis Jiménez-Linares**[1]

[1]Paseo de la Universidad n^o 4. Escuela Superior de Informática, Ciudad Real.
Universidad de Castilla-La Mancha.
{JavierAlonso.Albusac@uclm.es, JoseJesus.Castro, David.Vallejo, Luis.Jimenez}@uclm.es

Resumen

En los últimos años, el aumento generalizado de las medidas de seguridad, junto con la evolución tecnológica y el abaratamiento del hardware, han dado lugar a un incremento de interés por parte de la comunidad científica en el campo de la vigilancia inteligente. En este trabajo se describe cómo ha sido diseñado un componente de análisis de normalidad, basado en lógica difusa, para el estudio de relaciones espaciales entre los objetos móviles y zonas de una escena captadas por una cámara de vigilancia. Este componente será integrado dentro de un sistema de vigilancia cognitivo basado en técnicas de softcomputing, con una arquitectura dividida en varios niveles.

Palabras Clave: Vigilancia Inteligente, Lógica Difusa, Aprendizaje Inductivo.

1 INTRODUCCIÓN

Gran parte de los sistemas de seguridad implantados actualmente son una evolución de los sistemas de primera generación [3]. Dichos sistemas están formados por un conjunto de cámaras que envían la señal a una sede central donde un vigilante puede observar varios monitores. Según Tan Kok Kheng, vicepresidente de la división OEM de WPG Systems, - una de las principales distribuidoras de sistemas de vigilancia avanzados -, "Tras 20 minutos de vigilancia, la atención humana a los detalles del vídeo disminuye hasta niveles inaceptables y la videovigilancia deja de tener sentido. La videovigilancia tradicional ya no puede cumplir las, cada vez mayores, demandas del sector". La solución a estas deficiencias pasa por la utilización de sistemas de vigilancia inteligentes [3] capaces de interpretar lo que está ocurriendo. "La vídeovigilancia ya no puede ser simple y reactiva, necesita ser inteligente y proactiva", enfatiza el Sr. Kheng.

Entre las grandes dificultades a las que deben enfrentarse los Sistemas de Vigilancia Inteligentes (SVI) cabe destacar dos de ellas. En primer lugar, estos tipos de sistemas suelen observar subdominios del mundo real no deterministas, conviviendo en todo momento con la incertidumbre y la vaguedad, sin poder afirmar con total certeza qué clase de objetos están participando y qué eventos se están produciendo. Dependiendo de si se lleva a cabo una vigilancia forense o predictiva [5], existirá un mayor grado de incertidumbre o vaguedad. Cuando se realiza una vigilancia predictiva, el objetivo principal es anticiparse a posibles situaciones de riesgo antes de que ocurran, por lo que existe un alto grado de incertidumbre. En cambio, si la vigilancia es forense, es decir, se analizan los hechos una vez que han ocurrido, puede que la incertidumbre sea mínima y la vaguedad sea mayor.

Por otra parte, para que un SVI pueda clasificar un objeto a partir de sus propiedades o pueda determinar si las conductas y los eventos son normales, es necesario definir y formalizar el conocimiento del dominio. El segundo problema nace en la dificultad para reutilizar el conocimiento, debido en gran parte a la fuerte dependencia con el entorno observado. Por ejemplo, cada lugar observado por una cámara tiene sus zonas con sus características y normas propias. Un objeto observado se comportará de forma normal siempre y cuando cumpla dichas normas. Esta dificultad repercute directamente en el proceso de implantación o ampliación del sistema, ya que la ubicación de nuevos sensores implicaría la ampliación de la base de conocimiento por parte del experto. En este caso, los algoritmos de aprendizaje semi-automáticos y automáticos juegan un papel fundamental para agilizar este proceso y eliminar la fuerte dependencia con el experto [6, 2].

En el presente artículo se presentará un componente, basado en lógica difusa, integrado en un SVI en el que se está trabajando actualmente. Este componente estudia principalmente las relaciones espaciales que existen entre los objetos móviles y las zonas observadas, y clasifica los eventos simples a partir de un conjunto de reglas difusas generadas por un algoritmo de aprendizaje inductivo.

2 ARQUITECTURA DEL SVI

En la Figura 1 se puede observar la arquitectura multicapa del SVI con tres niveles bien diferenciados. En el primer nivel o nivel cero, se encuentran los sensores encargados de capturar el entorno y los algoritmos de visión que realizan la segmentación [7] y el tracking [8] de los objetos. En el segundo nivel se analizan los eventos que se producen en una escena y se determina si estos son normales o, por el contrario, implican algún peligro para el entorno.

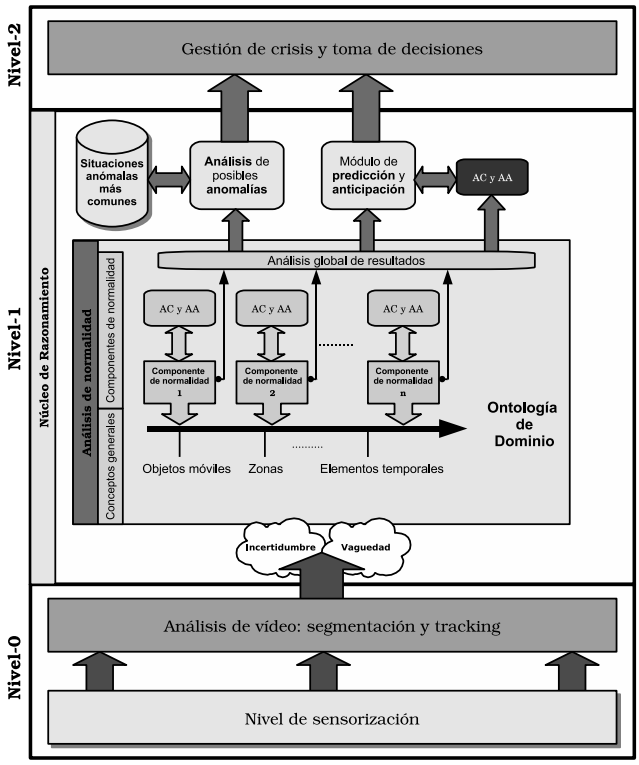

Figura 1: Arquitectura de un sistema de videovigilancia inteligente.

La mayoría de autores, para detectar situaciones anómalas, optan por definir las anomalías más comunes y tratan de encontrar similitudes entre éstas y las situaciones que se producen. El gran inconveniente de esta aproximación es que la mayoría de anomalías son imprevisibles incluso para un experto en el dominio. Por tanto, cuando se produce alguna de las situaciones anómalas no definidas, el sistema es incapaz de responder ante ellas. En cambio, la normalidad de un dominio es perfectamente conocida y, por este motivo, se ha optado por definir cuáles son las situaciones normales en un escenario, de tal forma que todo aquello que no sea normal debería ser tratado con especial atención. Además, para afrontar los problemas con mayor garantía de éxito, se definen las situaciones anormales más comunes para que sean identificadas. Por tanto, la peor situación se daría cuando el sistema fuera incapaz de recocer qué tipo de anomalía se está produciendo pero, al menos, sabría que está pasando algo que no es normal.

La formalización del conocimiento de dominio en el nivel intermedio se lleva a cabo mediante ontologías. En nuestro caso, la ontología de dominio está constituida, por una parte, por conceptos generales válidos para cualquier escena, como por ejemplo las clases de objetos móviles (personas, vehículos, etc) y sus propiedades, definición de zonas, elementos temporales, etc. Y por otra parte, los componentes de normalidad con módulos de adquisición y aprendizaje (AC y AA), que analizan la normalidad desde diferentes puntos de vista. Cada uno de estos componentes es independiente del resto y se pueden conectar o desconectar del sistema en cualquier momento. Así, el sistema puede por ejemplo disponer de un módulo encargado de analizar las trayectorias de los objetos móviles, otro componente para determinar si es apropiado que un tipo de objeto invada una determinada zona, otro componente para analizar las velocidades, etc. Cada componente ofrece respuestas en instantes de tiempo diferentes, dependiendo del peso del tipo de razonamiento. Uno de los requisitos fundamentales de cualquier sistema de seguridad es que ofrezca respuestas cercanas al tiempo real. Por tanto, son necesarios los componentes de normalidad ligeros, ya que a pesar de que no realizan un análisis exhaustivo de la situación actual, ofrecen una respuesta rápida.

Finalmente, el análisis de normalidad, la similitud con posibles situaciones anómalas y la predicción de acciones futuras, son los elementos que maneja el tercer nivel para tomar decisiones y gestionar una crisis en el caso de que ésta se produzca.

3 COMPONENTE DE NORMALIDAD DIFUSO

El objetivo principal del componente es la generación de reglas difusas que permitan clasificar los eventos de una escena, a partir de las relaciones espaciales que existen entre los objetos y las zonas. Se trata de un componente de normalidad ligero, ya que ofrece una

respuesta en un espacio breve de tiempo.

El motor de inferencia, a partir de las reglas obtenidas, tiene capacidad para determinar qué clases de objetos pueden invadir ciertas zonas y en qué grado (proporción de la superficie del objeto que invade la zona). Por ejemplo, una posible regla sería que un coche no puede invadir una zona ajardinada pero, si éste la invade parcialmente, el sistema podría decidir no activar la alarma ya que no sería un motivo suficiente. Por tanto, tratar las relaciones espaciales entre objetos y zonas de forma difusa, garantiza mayor flexibilidad al sistema y evita un número elevado de activaciones innecesarias de alarmas.

3.1 APRENDIZAJE INDUCTIVO DE REGLAS DIFUSAS

Sea ε un conjunto de ejemplos (eventos simples de una escena que ocurren en instantes de tiempo concretos) ε=$\{e_1, e_2, ..., e_n\}$, donde cada uno de ellos tiene la estructura $e_i = ((x_{i1}, x_{i2}, ..., x_{in}), o_i)$; siendo $x_{i1}, x_{i2}, ..., x_{in}$ los valores de entrada de las variables pertenecientes a V y o_i es la clase a la que pertenece el ejemplo.

El algoritmo de aprendizaje automático empleado por el componente utiliza como entrada el conjunto de ejemplos ε, las variables que describen cada ejemplo V, los dominios de definición de dichas variables DDV, y el conjunto de posibles salida [4]. Dicho algoritmo genera un conjunto de reglas difusas, que pueden ser fácilmente interpretadas, y cuya forma es la siguiente:

SI ν_1 es ZD_1 y ν_2 es ZD_2 y ... y ν_n es ZD_n **ENTONCES** clase = O_x,

donde cada ZD_i es un conjunto de valores disyuntivamente asociados con la variable ν_i, que toma valores de su dominio de definición DDV_i, verificando que $ZD_i \subseteq DDV_i$. Otra forma de escribir la regla es $((ZD_{1i}, ZD_{2i}, ..., ZD_{ni}), O_x)$.

El algoritmo se compone de dos pasos principales. El primero consiste en *"convertir los ejemplos de entrada en reglas particulares"*. Es decir, cada uno de los ejemplos del conjunto de entrenamiento se transforma en una regla, formada por la clase a la que pertenece el ejemplo y una etiqueta lingüística para cada variable (proceso de fusificación). El segundo paso consiste en *"Construir el conjunto de reglas definitivas maximales"* (utilizadas por el sistema para clasificar las situaciones) mediante un proceso de amplificación.

El principal problema en el proceso de clasificación es el de ambigüedad, que ocurre cuando un ejemplo de entrada cumple varias reglas pertenecientes a diferentes clases. En este caso, para resolver la ambigüedad, se elige aquella regla con mayor *grado de conveniencia*.

$$grado\_de\_conveniencia = min\{\varphi_i(x_j)\} \quad (1)$$

Donde φ_i son las funciones de pertenencia asociadas a cada variable de entrada X_j en la regla R_i. La funciones φ_i se construyen en función de las etiquetas presentes en la regla para cada X_j, por lo que éstas varían para cada una de las reglas.

3.2 VARIABLES DE DOMINIO

El componente de normalidad maneja un conjunto formado por cinco variables $V = \{v_1, v_2, v_3, v_4, v_5\}$, cada una de ellas con un dominio de definición $DDV = \{DDV_1, DDV_2, DDV_3, DDV_4, DDV_5\}$. Cada DDV_i de DDV se define mediante un conjunto $DDV_i = \{L_{1i}, L_{2i}, ..., L_{mi}\}$ que contiene todos los posibles valores que la variable v_i puede tomar (siendo L_{ji} la etiqueta lingüística del valor j en la variable v_i, definida mediante una función trapezoidal con los parámetros $(a_{ji}, b_{ji}, c_{ji}, d_{ji})$).

La variable v_1 representa la *clase* a la que pertenece un objeto observado (ver Fig. 2). El proceso de clasificación de un objeto se lleva a cabo en las capas inferiores del SVI durante el proceso de segmentación, liberando al componente de esta tarea. Cuando un objeto es clasificado, se le asigna un identificador que lo diferencia del resto y un valor numérico que representa la clase a la que pertenece. Finalmente, el componente de normalidad difuso sustituye el valor numérico por una etiqueta lingüística, necesaria para el algoritmo de aprendizaje.

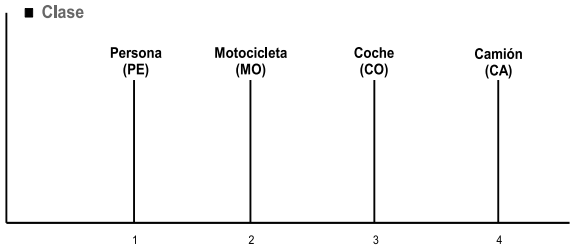

Figura 2: Variable v_1, clase a la que pertenece un objeto móvil. DDV_1={PE, MO, CO, CA}

La variable v_2 referencia el *tamaño* del objeto. El componente recibe de los niveles inferiores información acerca de la elipse que envuelve al objeto (el punto central y los dos radios cuya longitud se mide en píxeles). El valor lingüístico que representa el tamaño de un objeto se calcula a partir del área de la elipse que lo envuelve

$$\max_i\{\mu_i(\pi \times r_1 \times r_2)\} \mid i \in DDV_2$$

.

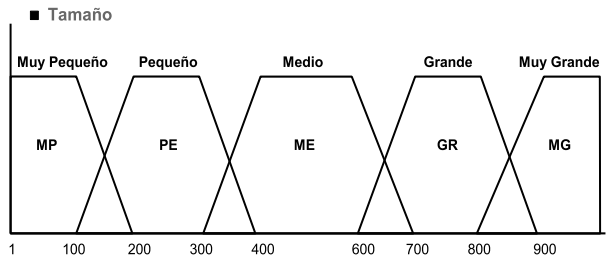

Figura 3: Variable v_2, tamaño de un objeto. DDV_2={MP, PE, ME, GR, MG}

Por otra parte, la variable v_3 representa la *intersección* entre un objeto y una zona (porcentaje de la superficie de un objeto que invade una zona). Para calcular el valor de dicha variable (Ec. 3), se recorre píxel a píxel el rectángulo mínimo que envuelve a la elipse de un objeto en la imagen capturada por la cámara (ver Fig. 3.2). En dicho recorrido se determina el número de puntos que pertenecen a la elipse (Ec. 2) y a la zona [1] en el rectángulo mínimo.

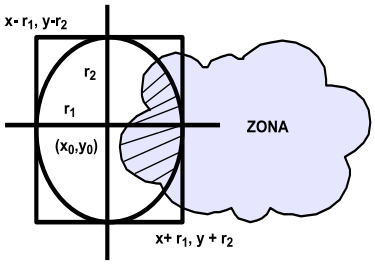

Figura 4: Rectángulo mínimo que envuelve a la elipse de un objeto e intersección con una zona

$$\frac{(x-x_0)^2}{r_1^2} + \frac{(y-y_0)^2}{r_2^2} \qquad (2)$$

Siendo (x_0,y_0) el centro de la elipse del objeto analizado y r_1, r_2, los radios de la elipse que representan el ancho y la altura respectivamente. La pertenencia del punto (x, y) a la elipse dependerá del valor obtenido en 2:

- < 1, el punto (x, y) pertenece a la elipse.
- $= 1$, el punto (x, y) pertenece al perímetro de la elipse.
- > 1, el punto (x, y) no pertenece a la elipse.

$$intersec(obj_x, z_y) = area(x \mid x \in obj_x \wedge x \in z_y) \qquad (3)$$

El último paso consiste en dar una interpretación lingüística de la intersección entre la elipse y la zona mediante el uso de DDV_3 (ver Fig. 5)

$$\max_i \{\mu_i(\frac{intersec(obj_x, zona_y)}{area(obj_x)})\} \mid i \in DDV_3$$

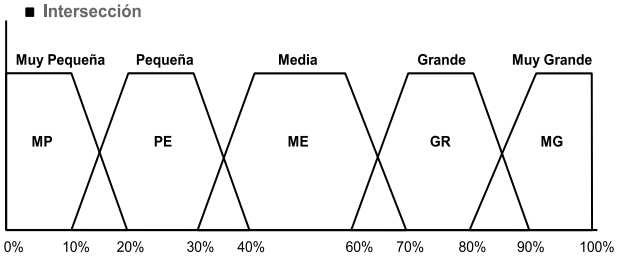

Figura 5: Variable v_3, tamaño de la intersección entre la elipse de un objeto y una zona. DDV_3={MP, PE, ME, GR, MG}

La variable v_4 representa la *velocidad* de un objeto móvil en una escena y depende del desplazamiento en píxeles desde un frame al siguiente. Por tanto, para determinar la velocidad de un objeto entre dos frames f_1 y f_2, se calcula la distancia (Ec. 4) entre los puntos (x_1, y_1) y (x_2, y_2) que corresponden con el centro de la elipse que envuelve al objeto en f_1 y f_2 respectivamente.

$$distancia(p_1, p_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \qquad (4)$$

Finalmente, se lleva a cabo una representación lingüística de los valores obtenidos a partir de la Ec. 4 con la utilización de DDV_4 (ver Fig. 6).

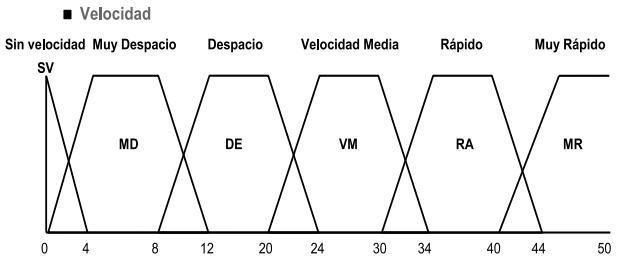

Figura 6: Variable v_4, velocidad de un objeto. DDV_4={SM, MD, DE, VM, RA, MR}

La última variable, v_5, representa la *zona* con la que intersecta el objeto. Un objeto puede invadir varias

zonas al mismo tiempo, por lo que la intersección con distintas zonas puede ser no nula simultáneamente. Cuando se da esta situación, el sistema trata cada una de las intersecciones de forma individual. De esta forma, si un objeto invade dos zonas al mismo tiempo se podría dar el caso en el que la relación espacial entre el objeto y una de las zonas fuera normal, mientras que con la otra podrían existir ciertas anomalías. Por otra parte, las zonas definidas en DDV_5 son totalmente dependientes de la escena observada, y en el caso de la Fig. 7 las zonas se corresponden con las mostradas en la Fig. 8 (ver Sección 4).

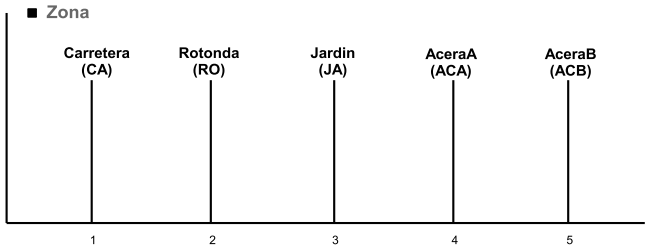

Figura 7: Variable v_5, zonas observadas en la escena. DDV_5={CA, RO, JA, ACA, ACB}

4 RESULTADOS EXPERIMENTALES

Para probar el componente de normalidad difuso se ha utilizado el vídeo capturado por una cámara IP situada en la ventana del Grupo de Investigación ORETO de la Universidad de Castilla-La Macha. La definición de los DDV_i para este ejemplo son los que se han mostrado en la seccción anterior. En la Figura 8, se muestran las zonas definidas para la escena observada. Dos aceras (Acera A y Acera B) por la que sólo pueden circular personas; la carretera por la que pueden circular tanto personas como vehículos; una zona ajardinada que separa dos carriles con sentidos opuestos (ningún objeto debería invadir esta zona); y finalmente, una rotonda que tampoco debería ser invadida por ningún objeto.

Observando la Figura 8, se puede intuir que la *perspectiva* puede ser una de las mayores fuentes de ruido, y por tanto, uno de los factores que pueden influir en mayor medida en los resultados. En la Figura 9 se pude apreciar como la elipse de un objeto, situado completamente en la carretera, invade parte del jardín.

La lógica difusa nos permite tratar con éxito este problema. El componente difuso y en concreto el algoritmo de aprendizaje, podría generar un conjunto de reglas que indiquen que los objetos con tamaño medio o grande, invadiendo el jardín muy poco o poco, es

Figura 8: Definición de zonas en la escena vigilada.

Figura 9: Problema de la perspectiva.

una situación normal. De esta forma, se elimina el ruido generado debido a los efectos ópticos producidos por la perspectiva. Por tanto, las reglas no tienen en cuenta la posición real y absoluta de un objeto en una escena, sino que analiza la posición de un objeto desde el punto de vista de la cámara.

A modo de ejemplo, en la Tabla 1 se muestran algunas de las reglas generadas a partir de un conjunto de ejemplos formado por 205 situaciones, de los cuales el 80% han sido empleados en el proceso de aprendizaje y el 20% para el proceso de clasificación. Ante la dificultad de obtener escenas de vídeo donde suceden situaciones anómalas, se ha optado por realizar simulaciones en las que se han ubicado objetos móviles virtuales en la escena.

En la Tabla 2 se muestran los resultados del proceso de aprendizaje y clasificación en diez tests diferentes. La mayoría de errores en el proceso de clasificación provienen de la ambigüedad que existe entre las reglas. En el contexto de los sistemas de seguridad, la peor situación ocurre cuando un evento anómalo es clasificado como normal. Por otra parte, si el algoritmo genera un número elevado de reglas, aumentan

Tabla 1: Reglas generadas de forma automática

Clase O_i	**Regla** R_i
O_1: Normal	R_0: Si v_3 no es {GR,MG} y v_5 no es {JA,ACA,ACB} R_1: Si v_3 no es {GR,MG} y v_4 no es {MR} y v_5 no es {ACB} R_2: Si v_2 no es {ME} y v_3 no es {ME,MG} y v_5 es {ACB,CA} R_3: Si v_2 no es {GR} y v_3 no es {ME,GR} y v_5 es {ACB}
O_2: Vehículo en acera	R_4: Si v_1 no es {PE} y v_2 no es {MP,PE} y v_4 es {GR,MG} y v_5 es {ACA,ACB}
O_3: Invasión del jardín	R_5: Si $v3$ no es {MP,PE} y v_5 es {JA}
O_4: Invasión de la rotonda	R_6: Si v_3 no es {MP,PE} y v_4 no es {MD} y v_5 es {RO} R_7: Si v_3 es {GR,MG} y v_5 es {RO}
O_5:Persona detenida en la carretera	Si v_1 es {PE} y v_2 es {MP,PE} y v_3 no es {MP,PE} y v_4 es {SV} y v_5 es {CA}

los conflictos entre ellas y, en consecuencia, la probabilidad de que el problema de la ambigüedad se acentúe. Por tanto, para reducir la ambigüedad en la mayor medida de lo posible, el experto debe configurar un conjunto de entrenamiento lo más representativo posible para así reducir el número de reglas y conflictos.

Tabla 2: Resultados de los tests

Test	**Reglas**	**Aciertos**	**Fallos**	**%**
1	17	39	1	97.5%
2	13	40	0	100%
3	15	37	3	92.5%
4	15	37	3	92.5%
5	13	38	2	95%
6	14	37	3	92.5%
7	15	38	2	95%
8	15	39	1	97.5%
9	16	36	4	90%
10	14	37	3	92.5%

5 CONCLUSIONES

En el presente artículo se ha presentado un componente difuso, integrado en un Sistema de Vigilancia Video Cognitivo, cuya función principal es clasificar los eventos que se producen en una escena observada por una cámara, mediante el análisis de las relaciones espaciales entre los objetos y las zonas del entorno.

Como línea de trabajo futuro, se pretende desarrollar otros componentes integrables que permitan analizar la normalidad desde otros puntos de vista. Además, se establecerá un mecanismo para relacionar la información procedente de diversos sensores heterogéneos ubicados en la misma escena, con el objetivo de reforzar la creencia y el porcentaje de éxito en las interpretaciones realizadas.

Agradecimientos

Este trabajo ha sido financiado gracias a los Proyectos PAC06-0141 y PBC06-0064 de la Junta de Comunidades de Castilla-La Mancha.

Referencias

[1] Shimrat, M. Algorithm 112, Position of Point Relative to Polygon. Comm. ACM 5(8), Pág.424, 1962.

[2] Makris, D. and Ellis, T. Automatic Learning of an activity-based semantic scene model. Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS'03), Pág.183-188, 2003.

[3] Valera, M. and Velastin, SA. Intelligent distributed surveillance systems: a review. IEEE Proceedings Vision, Image and Signal Processing, 152(2), Pág.192-204, 2005.

[4] Castro, JL., Castro-Schez, JJ and Zurita, J. Learning maximal structure rules in fuzzy logic for knowledge acquisition in expert systems. Fuzzy Sets and Systems, Elsevier, 101(3), Pág.331-342, 1999.

[5] Martínez-Tomás R., Ricón, M.,Bachiller, M. and Mira, J. On the correspondence between objects and events for the diagnosis of situations in visual surveillance task. Elsevier, Pattern Recognition Letters. 2007.

[6] Jiangungm, L., Qifeng, T., Tieniu, T. and Weiming, H. Semantic interpretation of object activities in a surveillance system. Proceedings of the 16 th International Conference on Pattern Recognition, ICPR'02. Pág.777-780, 2002.

[7] Rodríguez-Benítez, L., Moreno-García, J., Albusac, J., Castro-Schez, J.J. and Jiménez-Linares, L. An approximate reasoning technique for segmentation on compressed mpeg video. International Conference on Computer Vision Theory and Applications, VISAPP'07. Pág.184-191, 2007.

[8] Rodríguez-Benítez, L., Moreno-García, J., Albusac, J., Castro-Schez, J.J. and Jiménez-Linares, L. Seguimiento de objetos representados lingsticamente utilizando tcnicas de razonamiento aproximado. II Simposio sobre Lgica Fuzzy y Soft Computing (LFSC'07), II Congreso Espaol de Informtica, (CEDI'07). Pág.237-244, 2007.