



UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS  
VICERRECTORÍA DE FORMACIÓN  
DIRECCIÓN DE FORMACIÓN POSTGRADUADA

METODOLOGÍA PARA LA DETECCIÓN DE OBJETOS EN  
SISTEMAS PARA LA CATALOGACIÓN SEMI-AUTOMÁTICA  
Y AUTOMÁTICA DE VIDEOS.

*Tesis presentada en opción al título de Máster en Informática  
Aplicada*

Autor: **Ing. Yanio Hernandez Heredia**  
Tutor: **Dr. José Ortiz Rojas**  
Co-tutor: **Dr. Nicolás Guil Mata**

Ciudad de la Habana, septiembre de 2010

# DEDICATORIA

A mis padres y hermana, Jorge, Oraida y Yenía

A mi esposa Aylin

## DECLARACIÓN JURADA DE AUTORÍA Y AGRADECIMIENTOS

Yo Yanio Hernandez Heredia, con carné de identidad 83091024268, declaro que soy el autor principal del resultado que expongo en la presente memoria titulada “Modelación de una metodología para la detección de objetos en sistemas para la catalogación semi-automática y automática de videos”, para optar por el título de Máster en Informática Aplicada.

El presente trabajo fue desarrollado individualmente en el transcurso de los años 2009-2010

En especial deseo agradecer al Dr. José Ortiz Rojas de la UCI y a los Doctores. Nicolás Guil Mata y Julián Ramos Cozar de la Universidad de Málaga, España, quienes fueron tutor y co-tutores de mi investigación respectivamente y me ayudaron a la definición de esta investigación. Además, deseo agradecer a mis padres, hermana y mi bella esposa Aylin, los cuales me apoyaron en todo momento.

A todos ellos, así como a otros amigos que no he mencionado por razones de espacio, mi más sincero agradecimiento.

Finalmente declaro que todo lo anteriormente expuesto se ajusta a la verdad, y asumo la responsabilidad moral y jurídica que se derive de este juramento profesional.

Y para que así conste, firmo la presente declaración jurada de autoría en Ciudad de la Habana a los \_\_\_ días del mes de \_\_\_\_\_ del año 2010.

---

Firma del maestrante

## RESUMEN

Al observar un video o una imagen, es innato para el humano intentar hacer descomposiciones en distintos objetos o secciones. Cada uno de los objetos se puede interpretar dentro de la imagen por su estructura, textura, color, movimiento u otra característica. En el campo de la visión por computador, la detección de objetos es un paso crucial para sistemas de procesamiento de alto nivel, permite la cuantificación y visualización de los objetos de interés y de su calidad depende en gran medida el resultado final de la interpretación automática de la secuencias de imágenes que se analicen.

Son diversas las aplicaciones para la detección de objetos, dentro de las más importantes se encuentran la localización de rostros, logotipos o huellas en señales televisivas y la detección de lesiones y/o enfermedades sobre imágenes médicas. Existen diversas técnicas para la extracción y clasificación de características a fin de detectar objetos en secuencias de imágenes; se encuentran las basadas en la matemática, sobre todo con fines médicos: diferencias gaussianas, redes neuronales, morfología matemática, Fourier y Wavelets, de igual forma las imágenes que contienen otro tipo de objetos se han analizado con técnicas que se basan en la información de color y textura. Es conocido que hasta ahora, no existe un método universal de detección ya que cada método está íntimamente ligado con el problema específico a resolver, es por esto que se hace necesario desarrollar una metodología que permita detectar diferentes clases de objetos en imágenes digitales.

En este trabajo se propone una metodología que utiliza técnicas híbridas de procesamiento de imágenes para la detección de objetos, teniendo en cuenta que permiten, aprovechando las ventajas de cada cual, la extracción de características discriminantes que representen a cada uno de los objetos existentes en una imagen. La metodología incluye en primer lugar la segmentación de los videos a analizar, lográndose de esta forma varios hilos de procesamiento simultáneos, luego se extraen los fotogramas más representativos para disminuir el costo computacional de las comparaciones con los objetos almacenados y creando

clasificadores débiles con las características no compatibles de los objetos desechados, para que en el futuro se conviertan en objetos descriptivos.

Los resultados obtenidos con pruebas de ensayo realizadas en MATLAB y C++ con OpenCV, son bastantes alentadoras y demuestran la posibilidad de usar esta metodología en entornos reales de aplicación, con un alto grado de precisión y usabilidad. Encuestas aplicadas a un grupo de expertos, respaldan la necesidad y posibilidad de la propuesta realizada, así como la novedad de la investigación.

## ÍNDICE GENERAL

INTRODUCCIÓN .....	1
Antecedentes .....	2
Estructura del Documento.....	4
Glosario de Términos .....	5
CAPÍTULO 1 .....	6
Fundamentación Teórica de la investigación .....	6
1.1 Conceptos asociados al dominio del problema .....	6
1.2 Técnicas para el reconocimiento de Objetos .....	11
1.3 Resultados de investigaciones asociadas.....	20
1.4 Conclusiones parciales .....	24
CAPITULO 2 .....	26
Diseño de la metodología.....	26
2.1 Estructura de la Metodología .....	26
2.2 Descripción de los componentes de la metodología .....	27
2.3 Conclusiones Parciales .....	35
CAPÍTULO 3 .....	37
Análisis de los resultados.....	37
3.1. Herramientas de ensayo .....	37
3.2. Experimentos .....	39
3.3. Encuestas a expertos.....	43
3.4. Conclusiones Parciales .....	47
Conclusiones.....	49
Recomendaciones.....	50
Publicaciones derivadas.....	51
Referencias bibliográficas .....	52
Anexos .....	59

## ÍNDICE DE FIGURAS

Figura 1. Segmentación Temporal de Video .....	7
Figura 2. Fotogramas claves en una secuencia de video segmentada .....	8
Figura 3. Histograma de una imagen. ....	10
Figura 4. Detección de regiones de brillo como objetos .....	12
Figura 5. Detección de objetos por segmentación multinivel .....	12
Figura 6. Regiones de las características discriminatorias en objetos. ....	14
Figura 7. a) Detección de contornos de imagen segmentada RGB, b) Detección de contornos de imagen segmentada por combinación RGB-HSV.....	20
Figura 8. Curvas Precision- recall para dos tareas de detección de objetos en la competición PASCAL VOC 2005 Challenge .....	21
Figura 9. Curvas Precision / Recall para modelos entrenados con las categorías carros y personas de la base de datos de la competencia PASCAL 2006.....	22
Figura 10. Resultados de la detección en las Bases de Datos TUD Peatones y UIUC-Escala.....	23
Figura 11. Diagrama de la Metodología propuesta .....	27
Figura 12. Algoritmo de detección de los puntos de alta curvatura .....	32
Figura 13. Gráfica con los resultados de los bosques de Hough .....	41
Figura 14. Tabla con los valores de precisión y recal.....	42
Figura 15. Grafica Precission – Recall .....	42
Figura 16. Aplicación General para tareas de análisis e indexación de videos .....	59
Figura 17. Comportamiento de diferentes métodos en las dos Bases de Datos de coche de UIUC .....	60
Figura 18. Diagrama de clases Forest Detector. Implementación en C++, compilada para Linux con OpenCV .....	61
Figura 19. a) Weizzman Horse DB    b) Airplanes Dates Trevick DB (Imagen de Hough, Gráfica de intensidad en el punto medio, imagen original con el Bounding box). ....	62

## INTRODUCCIÓN

La detección automática de información y descriptores con que cuentan los videos en diferentes campos de aplicación (D. Hoiem, 2008), ha sido uno de los temas más estudiados en los últimos años por diferentes grupos multidisciplinarios en el mundo que se dedican a la investigación del procesamiento de imágenes y señales de video (P. Gil, 2004). El incremento constante de las capacidades de almacenamiento en los sistemas de cómputo ha permitido crear grandes volúmenes de datos de materiales de audio y video, a los cuales se les puede aplicar de forma digital y en diferentes formatos, técnicas para extraer información indexada en los mismos. Estas técnicas o algoritmos para realizar análisis automáticos o semi-automáticos requieren de gran eficiencia en sus resultados, buscando obtener dos características fundamentales: rapidez y precisión en los resultados alcanzados.

El procesamiento automático de imagen, se desarrolla aceleradamente en los últimos años, sobre todo por la gran cantidad de aplicaciones que cuentan con: segmentación automática, comparación entre secuencias de video almacenadas y nuevas, detección de firmas, logos y objetos. No obstante, según la bibliografía consultada, es posible afirmar que ninguna técnica desarrollada para estos fines ha tenido un porcentaje de fiabilidad elevado, sobre todo por dos razones: diferentes características que atentan contra la variación de tipos de videos a analizar y por otro lado, la necesidad de implementarle a los algoritmos de detección, una funcionalidad conocida como *Machine Learning* (Escolano, 2007). Las pruebas se realizan con un cierto número de muestras de videos y elementos a reconocer, contenidos en base de datos, pero es necesario que con el paso del tiempo, estas muestras se incrementen de forma automática, en el peor de los casos de manera semi-automática.

La detección de objetos de cualquier naturaleza es una área en la que se han obtenido pocos resultados dentro de la visión artificial por computador, sin embargo son elevados los esfuerzos que se realizan para desarrollarla por la gran aplicación que tiene en la gestión de archivos audiovisuales, con la clasificación

detallada de materiales, basada en descriptores que tributen al reconocimiento de variados objetos. La utilización e integración de técnicas existentes para el procesamiento de secuencias de imágenes, ha sido también una variante que han encontrado algunos desarrolladores para obtener buenos resultados (Laptev, 2005).

## **Antecedentes**

El desarrollo del trabajo surge por la creación de un producto informático para la captura y catalogación de medias. Este software para poder competir en el mercado internacional debe constar con funcionalidades esenciales como: captura de señales de video según planificaciones definidas, fácil cambio de la señal que se está capturando en cada momento, almacenamiento de los ficheros de video capturados en un archivo, grabación de señales de audio o radio FM, visualización de los materiales audiovisuales almacenados, catalogación de las medias almacenados, transcripción de ficheros de audio, obtención de reportes predeterminados y configurables en el sistema.

Obtener un producto informático con todas estas prestaciones implica un gran impacto en el mercado nacional e internacional. Sin embargo en la captura o grabación de medias, específicamente videos, se aplican diversas técnicas y algoritmos matemáticos para la extracción automática de metadatos relacionados con características de las medias y una agilización en el proceso de catalogación que pueden añadir al producto valores agregados. A pesar de no definirse para la primera versión del producto diversas funcionalidades de detección automática, sí fue preciso el estudio del arte de los sistemas de análisis de video inteligentes, obteniéndose resultados en la segmentación automática y extracción de fotogramas claves.

Dada la importancia de la detección automática de objetos en videos e imágenes no solo en sistemas de esta naturaleza, sino de manera general para muchos otros de ciencias que integran el tratamiento, análisis e interpretación de imágenes con fines específicos, como por ejemplo la salud y la seguridad vehicular, por solo citar algunos, se plantea como ***problema de la investigación***,

la necesidad de diseñar una metodología que utilice técnicas eficientes para la detección de objetos aplicadas a los sistemas para la catalogación semi-automática y automática de información audiovisual.

### ***Objeto de Estudio***

El objeto en el cual se enmarca el desarrollo de este trabajo, desde el punto de vista práctico y teórico para darle solución al problema planteado, consiste en el estudio de algunas de las diversas técnicas y algoritmos matemáticos más conocidos que existen para la detección de objetos en videos, así como evaluaciones en términos de fiabilidad de reconocimiento.

### ***Campo de Acción***

El campo de acción de esta investigación lo constituye la búsqueda y detección de objetos de cualquier naturaleza en secuencias de video de forma automática, en aplicaciones de catalogación de medias.

### ***Objetivo General***

Con el fin de plantear una solución al problema descrito, se propone como objetivo general diseñar una metodología que permita detectar objetos en sistemas de catalogación semi-automática y automática de videos, teniendo en cuenta el estudio y aplicación de diversas técnicas existentes en la literatura que se utilizan para resolver problemas similares de manera eficiente.

### ***Hipótesis***

Disímiles son las técnicas matemáticas y algoritmos que se aplican para darle solución al problema descrito o algunos similares. Con la siguiente investigación, se propone una variante de solución partiendo de la siguiente hipótesis:

Si se modela una metodología eficiente para el reconocimiento de objetos en videos e imágenes, se obtendrá una herramienta aplicada a la catalogación automática, para minimizar el tiempo de búsqueda y comparación en las base de datos de almacenamiento de objetos, así como el costo computacional utilizado para el procesamiento de dichas tareas de reconocimiento.

## ***Tareas de Investigación***

Como tareas de investigación se proponen las siguientes:

- Identificar las variantes de técnicas y algoritmos existentes, así como las tendencias utilizadas en la solución del problema planteado, a partir de un estudio del estado del arte.
- Seleccionar las técnicas eficientes de búsqueda y comparación de objetos almacenados en secuencias de video a utilizar en el diseño de la Metodología propuesta.
- Diseñar una Metodología para detectar objetos en secuencias de video utilizando las técnicas seleccionadas.
- Validar la Metodología propuesta con la aplicación del método Delphi como herramienta de investigación multidisciplinaria para la realización de pronósticos y predicciones y pruebas realizadas a las técnicas que incluye la metodología.

## **Estructura del Documento**

El presente documento se encuentra dividido en tres capítulos. En el primero de ellos, Fundamentación Teórica, se realiza un análisis de algunas de las técnicas más eficientes para reconocer objetos descritas en la literatura, además se exponen elementos esenciales sobre el procesamiento de videos digitales y el origen y descripción de metodologías como concepto.

El segundo capítulo, describe los componentes fundamentales referentes a la metodología propuesta y los elementos que incluye la misma. Comienza exponiendo un conjunto de conceptos referentes a la definición de sus componentes, continúa representando la relación existente entre las técnicas propuestas, y culmina con una descripción del flujo definido en la modelación de la metodología.

En el tercer capítulo se presentan los resultados obtenidos a través de pruebas realizadas a las técnicas que se utilizaron en la concepción de la

metodología. Para el desarrollo de las pruebas se seleccionaron bases de datos de objetos utilizadas en varias competiciones de visión por computador a nivel mundial (Everingham, 2007) (L. Van Gool, 2006). Todo el análisis realizado en este último capítulo permitió obtener importantes conclusiones y recomendaciones para la futura aplicación de la metodología.

## **Glosario de Términos**

**Segmentación:** Dividir una imagen en regiones u objetos.

**Clustering:** Puede ser considerado el más importante problema de aprendizaje no supervisado, de modo que, como todos los demás problemas de este tipo, se trata de encontrar una estructura en un conjunto de datos no etiquetados (Zaiane, 2003).

**Frame:** Se denomina **frame** (en español, *fotograma*) a cada una de las imágenes que conforman un video o impresas en un papel.

**Shot:** Son los diferentes cambios de toma que pueden existir en un video.

**GPU:** es una unidad de procesamiento gráfico (también a veces es llamada unidad de procesamiento visual o VPU) especializada, que descarga la representación de los gráficos en 2D y 3D del microprocesador.

# CAPÍTULO 1

## Fundamentación Teórica de la investigación

Con el objetivo de facilitar la comprensión del alcance de la investigación, en el presente capítulo se exponen conceptos fundamentales asociados al dominio del problema planteado. Se realiza un análisis del estado del arte que precede a la realización de este trabajo y que contribuye a esclarecer su objeto de estudio. Se presentan las principales técnicas existentes para el reconocimiento de objetos en secuencias de imágenes, así como los resultados que se han obtenido con su utilización, lo que contribuye a la selección de las más adecuadas a tener en cuenta para la composición de la Metodología a proponer.

### Conceptos asociados al dominio del problema

A continuación se relacionan los principales conceptos que ayudan a entender el desarrollo de la investigación.

#### **1.1.1. Metodología**

La Metodología, hace referencia al conjunto de procedimientos basados en principios lógicos, utilizados para alcanzar una gama de objetivos que rigen una investigación científica. Un metodólogo no pone en tela de juicio el conocimiento ya obtenido y aceptado por la Comunidad científica. Su problema se centra en la búsqueda de estrategias válidas para incrementar el conocimiento. Por ello, la metodología es parte del proceso de investigación o «Método Científico», que permite sistematizar los métodos y las técnicas necesarias para llevarlas a cabo. Los métodos elegidos por el investigador facilitan el descubrimiento de conocimientos seguros y confiables que, potencialmente, solucionarán los problemas planteados en un proyecto de investigación.

#### **1.1.2. Segmentación**

El análisis de secuencias de video orientado a la extracción automática de información referente a su contenido es un proceso genéricamente conocido como video parsing (Akutsu, 1992).

El primer paso de dicho análisis consiste habitualmente en llevar a cabo una segmentación temporal de la secuencia de video, es decir, una subdivisión o estructuración en unidades homogéneas desde algún punto de vista (ver Figura 1. Segmentación Temporal de Video), ya sea objetivo (luminosidad media, distribución de color, movimiento de la cámara, entre otras características) o subjetivo (coherencia de contenido) (Brunelli, 1997). En primer lugar, ello permite disponer de un índice de la secuencia que posibilita un acceso eficaz a partes de la misma. En segundo lugar, la homogeneidad de las subsecuencias obtenidas facilita la aplicación de técnicas de extracción manual o automática de características (personajes, marcas publicitarias, letreros, primeros planos o paisajes) orientadas a anotar los contenidos de la secuencia, de ahí que se conozcan como técnicas de catalogación de video.

*Definición de segmentación:* Proceso al cual es sometido un fichero audiovisual digital para ser dividido en partes, teniendo en cuenta criterios de selección para proceder a su posterior análisis con mayor facilidad (Cheriet, 1998).



Figura 1. Segmentación Temporal de Video

### 1.1.3. Fotograma clave

La definición de un fotograma clave parte del propio fotograma, no es más que un espacio en el cual podemos obtener objetos vectoriales, imágenes o símbolos; un conjunto de fotogramas conforman una animación o video como se muestra en

Figura 2. Fotogramas claves en una secuencia de video segmentada (Macromedia, 2000).

Los fotogramas clave son fotogramas de videos completos (o imágenes) insertados a intervalos concretos en un clip de video. Los fotogramas que están entre fotogramas clave contienen información sobre el movimiento y los cambios de escena que tienen lugar entre dichos fotogramas clave. Por ejemplo, si un video muestra una persona que pasa por una puerta, los fotogramas clave contienen la imagen completa de la persona y de la puerta en el fondo, y los fotogramas intermedios incluyen la información que describe el movimiento de la persona conforme pasa por la puerta.

De forma predeterminada, algunos sistemas calculan automáticamente el intervalo para insertar los fotogramas clave en función de la velocidad de fotogramas del clip de video. El valor del intervalo de fotogramas clave indica al codificador la frecuencia con que debe volver a evaluar la imagen de video y grabar un fotograma completo o fotograma clave (Resource, 1999).



Figura 2. Fotogramas claves en una secuencia de video segmentada

#### **1.1.4. Histograma bidimensional**

Se define histograma de una imagen como la curva que en ordenadas representa cada uno de los posibles niveles de gris (0 - L), donde L es el máximo nivel de gris - 255, mostrando en abscisas la frecuencia relativa de aparición del

mismo en la imagen (Figura 3) (Starópoli, 2006). También se puede definir como un gráfico que muestra la distribución de los colores o tonos de un color en una imagen según su luminosidad.

El histograma bidimensional es útil, sobre todo cuando se tienen imágenes multiespectrales. El más común es cuando se tiene imágenes de color descompuestas en sus tres canales: RGB (componentes roja, verde, y azul) (Campos, 1999). Este histograma contiene dos variables cartesianas, luminancia y saturación. Su finalidad, radica en que proporciona un mecanismo sencillo e intuitivo para separar la información cromática de la acromática. La evaluación de un histograma bidimensional y la fijación de umbrales sobre él, permiten conseguir la extracción de reflejo y de distintos grados de sombras (Angulo, 2003).

Para utilizar los histogramas en la visión artificial es importante conocer algunas de sus propiedades: cuando una imagen se condensa o se representa por su histograma, se pierde toda la información espacial, dada una imagen, su histograma es único, pero el inverso no es verdadero. Esto quiere decir que imágenes diferentes, pueden tener el mismo histograma (Starópoli, 2006). No obstante es difícil que ocurra en una misma secuencia de video.

En el campo del procesamiento de imágenes, la elección del umbral para la frontera es determinante en muchos algoritmos. Al método de usar líneas de contorno como fronteras, se llama técnica de umbral (del inglés thresholding). A partir del histograma se puede encontrar el umbral óptimo para la determinación de una frontera, sin necesidad de ver la imagen.

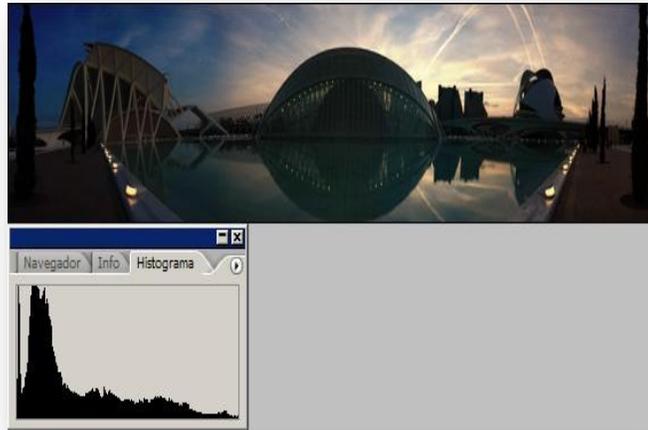


Figura 3. Histograma de una imagen.

### 1.1.5. Machine Learning

El objetivo de los métodos “Machine learning” es programar las computadoras para uso de datos de ejemplo o la experiencia para resolver un problema dado. Muchas aplicaciones con éxito del aprendizaje de la máquina ya existen, optimizar el comportamiento para que una tarea pueda ser completada con recursos mínimos, y extraer conocimiento de los datos existentes, es la premisa fundamental. Son varios los algoritmos de aprendizaje que aparecen en la web y que se explican de manera que pueda llevarse, de las ecuaciones a un programa de ordenador (MIT, 2004).

Esta funcionalidad no es solamente utilizada en trabajos de clasificación de selectores. Las siguientes clases con que cuenta, trabajan los principales problemas que existen asociados al campo:

- Clasificación de aprendizaje: aprender a poner variables de casos en clases predefinidas.
- Asociación de aprendizaje: aprender las relaciones entre los atributos que pertenezcan a una clase.
- Clustering: descubrir las instancias de clases que conforman una asociación de objetos.

- Predicción numérica: aprender a predecir una cantidad numérica que se instancie en una clase determinada (Keller, 2006).

## **Técnicas para el reconocimiento de Objetos**

En este epígrafe se relacionan algunas técnicas, para el reconocimiento de objetos, con resultados internacionales que servirán de base para la fundamentación investigativa de la propuesta en el capítulo 2.

### ***1.2.1. Detección de objetos por segmentación multinivel combinada de espacios de color***

El método expuesto en este epígrafe utiliza la información de color tratando de reducir la influencia de sombras y brillos que aparecen, formando parte de información acromática presente en las imágenes capturadas. Para ello, se hace uso de las representaciones RGB y HSV, y de una aproximación multiumbral del método de Otsu<sup>1</sup> (Otsu, 1979).

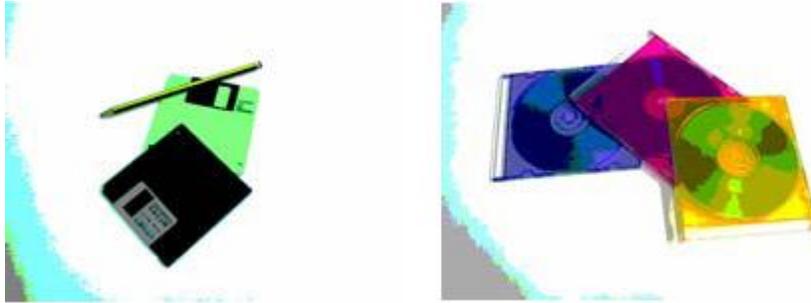
El método de detección, que aquí se muestra, se compone de dos fases, por un lado la segmentación en regiones cromáticas y por otro lado la detección de los contornos de esas regiones (P. Gil, 2004).

La segmentación de una imagen RGB<sup>2</sup> se realiza segmentando individualmente cada una de las tres componentes de color de las que consta la imagen. Así, se aplica el método de Otsu a cada una (Yingzi, 2004), previamente suavizadas para provocar mejoras en el histograma que permitan obtener los umbrales con un mayor éxito. La ventaja que ofrece el método de segmentación multi-umbral RGB se basa principalmente en que es capaz de detectar información cromática con tonos de color muy diversos incluyendo los conocidos acromáticos (escala de grises). Sin embargo, también, obliga a que se detecten como objetos regiones de sombras y brillos, que en esta segmentación serán considerados como objetos de un determinado color grisáceo, Figura 4.

---

<sup>1</sup> calcula el valor umbral de forma que la dispersión dentro de cada segmento sea lo más pequeña posible, pero al mismo tiempo la dispersión sea lo más alta posible entre segmentos diferentes.

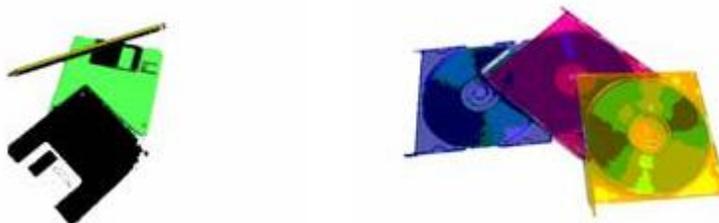
<sup>2</sup> Es un modelo de color aditivo en la que el rojo, verde y azul se suman en varias formas para reproducir una amplia gama de colores.



**Figura 4. Detección de regiones de brillo como objetos**

Por otro lado, el método multi-umbral HSV<sup>3</sup> (Alfredo, 2008) permite detectar objetos de color sin detectar por ello regiones de sombras y brillos que al ser acromáticas, generalmente, son desechadas como objetos válidos. Como inconveniente, destacar que un objeto en escala de grises no será detectado correctamente, al considerarse o bien un brillo o bien una sombra.

Vistos los problemas que acarrearán uno y otro método de segmentación, se ha optado en esta propuesta por probar una combinación de ambos. De modo que inicialmente, se busca detectar los objetos presentes en la imagen haciendo uso de la segmentación RGB y posteriormente, haciendo uso de algunas de las informaciones que proporcionan las componentes del espacio de color HSV de la imagen original, realizar una post-segmentación de la imagen segmentada mediante RGB, como se muestra en la Figura 5. Así, se busca detectar todos los objetos en la imagen incluidos los que carecen de cromaticidad y por otro lado, eliminar en la medida de lo posible los brillos y las sombras, buscando conseguir homogenización de las áreas detectadas.



**Figura 5. Detección de objetos por segmentación multinivel combinada de espacios de color**

<sup>3</sup> Cada color trabaja con 3 componentes básicas: matiz, saturación y brillo

### 1.2.2. Detección de objetos con boosted histogramas

En esta técnica, similar a la utilizada por (K. Levi, 2004) se selecciona la posición y la forma de las características del histograma para minimizar el error de entrenamiento para una determinada tarea de reconocimiento. Durante el entrenamiento, se considera que una exhaustiva cantidad de regiones rectangulares en la ventana de objeto normalizado y se calculan los descriptores del histograma para cada uno de ellos utilizando de manera eficiente los histogramas integrales<sup>4</sup> (Porikli, 2005).

Luego se aplica el algoritmo AdaBoost (P. Viola, 2001), para seleccionar características del histograma y para que aprenda el clasificador de objetos. La gran contribución de (Laptev, 2009), en cuanto a aprendizaje se refiere, es la adaptación del Framework Boosting, al vector evaluado con las características del histograma y el diseño de un “weak learning”<sup>5</sup> ponderado sobre la base del Discriminante Lineal de Fischer (WFLD de sus siglas en inglés). Además de desplegar la posición dependiente de las características del histograma y artificialmente se aumenta el tamaño del conjunto de entrenamiento, mediante la adición de ruido espacial de la anotación. Estas extensiones muestran una sustancial mejora respecto a la definición propuesta por (K. Levi, 2004).

AdaBoost (Y. Freund, 1997) usado en este algoritmo propuesto, es una popular máquina de aprendizaje que tiene como método de funcionamiento, la combinación de propiedades de un clasificador eficaz y la selección de características. La versión discreta de AdaBoost define un fuerte Clasificador binario  $H$ , que se define como sigue:

$$H(z) = \text{sgn} \left( \sum_{t=1}^T h_t(z) \right)$$

Durante el entrenamiento se asume una alineación aproximada de las muestras de objetos con una ventana rectangular como se puede apreciar en la

---

<sup>4</sup> Método que permite emplear un proceso de búsqueda exhaustiva en tiempo real, que puede llegar a ser práctico

<sup>5</sup> Se le nombra de esta manera por la adaptabilidad en el aprendizaje basado en características no predominantes

Figura 6. Bajo este supuesto se basa (Laptev, 2009) en la correspondencia de las partes objeto y se aprende de la apariencia de las partes de las regiones de imágenes correspondientes. Para evitar una selección heurística de esas regiones, inicialmente se consideran como un conjunto exhaustivo de sub-ventanas rectangulares en el objeto, las potenciadas en el aprendizaje con Ada.

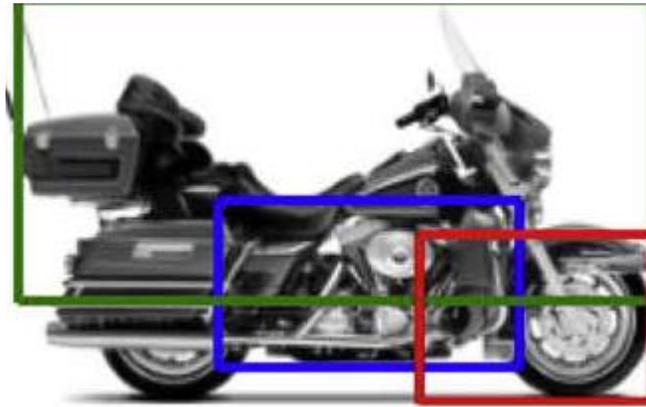


Figura 6. Regiones seleccionadas donde están la mayoría de las características discriminatorias.

Luego se representa cada característica por un histograma de una imagen local con sus regiones y mediciones. Se adoptan las características de los Histogramas orientados del gradiente (HOG por sus siglas en inglés) y se consideran como alternativa, los histogramas de las mediciones de las imágenes, con la segunda derivada para la imagen y el color.

### ***1.2.3. Aplicación de Redes Neuronales para el reconocimiento de Objetos***

Este método de detección de objetos, propone implementar una Red Neuronal Perceptrón<sup>6</sup> Multicapa con la librería ml (Machine Learning) del OpenCV<sup>7</sup> para poder realizar el reconocimiento de objetos. Se pueden considerar tipos de objetos con formas definidas para realizar el entrenamiento. Estas formas para el ejemplo utilizado en (Aquiye, 2009) han sido, circulares, cuadráticas, triángulos y estrellas.

---

<sup>6</sup> Usa una matriz para representar las redes neuronales y es un discriminador terciario que traza su entrada  $x$  (un vector binario) a un único valor de salida  $f(x)$  (un solo valor binario) a través de dicha matriz.

<sup>7</sup> Biblioteca libre de visión artificial originalmente desarrollada por Intel y publicada bajo licencia BSD

Además se utilizaron imágenes a colores, que fueron procesadas para obtener imágenes segmentadas en dos clases (binarizadas) a partir de las cuales se calcularon las regiones para realizar el entrenamiento prueba de la red neuronal.

Para el reconocimiento, las imágenes capturadas son transformadas a escala de grises y luego se binarizan considerando un umbral cuyo valor es 115, el cual fue definido como el promedio entre el mayor y menor valor del histograma. Mediante las marcas, obtenidas de las librerías que ya están definidas para el OpenCV, se pueden seleccionar las regiones donde están los objetos que se encuentran en las imágenes binarizadas.

Para el entrenamiento y prueba de la Red, se utilizó la Red Perceptrón Multicapa como arquitectura. Una vez obtenidas las imágenes binarizadas se entrena la red con ellas, definiendo las entradas, como una matriz de NUM x 3, donde NUM es la cantidad de datos ingresados para el entrenamiento de la Red, y 3 son las capas de entrada consideradas como área, perímetro y compacidad. Entre las capas de entrada y de salida, hay 2 capas ocultas definidas cada una de 13, que generan como salida las cuatro capas que definen las características para el entrenamiento, los cuales son: área, perímetro y compacidad generando una matriz de salida de NUM x 4 (Aquiye, 2009).

#### ***1.2.4. Detección de objetos a partir de modelos multi escalas de partes deformables***

Se describe un sistema de detección de objetos sobre la base de mezclas de los modelos multi escalas de partes deformables. El sistema es capaz de representar a clases de objetos muy variables, que se presentan en las competiciones de gran prestigio “PASCAL Visual Object Classes”, (Everingham, 2007). Este algoritmo utiliza métodos para el entrenamiento discriminativo con datos etiquetados y se logra combinar un acercamiento sensitivo para los

ejemplos negativos duros de minerías de datos con un formalismo que se denomina SVM<sup>8</sup> latente (Felzenszwalb, 2008).

Este enfoque se basa en un framework de estructuras pictóricas (Elschlager, 2005), donde los objetos se describen por un conjunto de elementos organizados en una configuración deformable. En una estructura pictórica el modelo de cada parte codifica las propiedades de apariencia local de un objeto, y la deformable se caracteriza por conexiones entre ciertos pares de elementos.

Sin embargo, si bien estos modelos son atractivos desde el punto de vista conceptual, ha sido difícil establecer su valor en la práctica, mayormente son superados por modelos más simples, tales como plantillas rígidas (Triggs, 2005).

La primera innovación de este algoritmo consiste en enriquecer el modelo de Dalal-Triggs (Triggs, 2005), utilizando una estrella estructurada a partir del modelo definido por la raíz de un "filtro" (análogo al del algoritmo antes mencionado), además de una colección de una parte de los filtros y los modelos asociados a la deformación.

Por otro lado, se utiliza para entrenar el modelo una máquina de soporte vectorial latente (LMSV) (Stuart Andrews, 2003), donde cada ejemplo es marcado por la siguiente función:

$$f\beta x = \max_{z \in Z_x} \beta * \varphi(x, y)$$

Donde  $\beta$  es un vector con los parámetros del modelo,  $z$  los valores latentes y  $\varphi(x, y)$  es el vector característico. En el caso del algoritmo que proponen Felzenszwalb y Girshick sería la concatenación del filtro de la raíz, las partes del filtro y los pesos del costo de deformación.

Existe una gran variedad de modelos deformables para la detección de objetos, que incluyen además modelos de plantillas deformables (Cootes, 2003) (Coughlan, 2000) (Huttenlocher, 2004) o variedades del modelo basado en partes

---

<sup>8</sup> Máquina de soporte vectorial: es un tipo de máquina de aprendizaje que se entrena con varios casos reales con respuestas para que en un futuro logre tomar decisiones

(R. Fergus, 2004) (B. Leibe, 2008), los ejemplos señalados demuestran la gran utilización en la actualidad de estos algoritmos en la visión por computador, específicamente en la detección de objetos.

#### ***1.2.5. Bosques de Hough para la detección de objetos de clases específicas***

Se presenta un método para la detección de objetos de una clase específica, tales como automóviles o peatones, en imágenes naturales. Esto se logra a través de la generalizada transformada de Hough muy utilizada para estos fines en el campo de la visión por computador, donde las detecciones de los distintos parches de un objeto permiten emitir votos probabilísticos para posibles ubicaciones del centroide de todo el objeto, la detección de la hipótesis corresponde a la máxima de la imagen de Hough que acumula los votos de todas las partes.

Sin embargo, (Lempitsky, 2009) considera que los métodos anteriores detectan las partes de los objetos utilizando un código generativo de las apariencias de las partes, tomaron un enfoque más discriminativo que consiste en formar un bosque de Hough para clases específicas de imágenes, que no es más que un conjunto de árboles aleatorios que directamente asignan a los parches de la imagen la votación sobre las probabilidades de la posible ubicación del centroide del objeto.

Los bosques aleatorios han atraído recientemente mucha atención en la visión por computador (F. Schroff, 2008) (R. Cipolla, 2008) (Shotton, 2007). Relacionado a la técnica propuesta por (Lempitsky, 2009), la idea de reemplazar los libros de códigos generativos por un conjunto de árboles aleatorios se ha investigado en el contexto de la clasificación y segmentación semántica de las imágenes.

Mientras que los bosques de Hough son en muchos aspectos similares a otros bosques aleatorios en visión por computador, poseen interesantes propiedades específicas, motivados por su uso dentro de la transformada de Hough generalizada:

- Los árboles de un bosque Hough se construyen directamente para optimizar el rendimiento de los votos. La formación se obtiene de cada árbol para que las hojas produzcan votos probabilísticos a favor de obtener una pequeña incertidumbre.
- Cada árbol se construye sobre la base de la colección de parches extraídos de datos de entrenamiento. Es importante destacar que el proceso emplea todos los datos disponibles para la supervisión de entrenamiento, si uno viene de un parche de fondo o un objeto (Lempitsky, 2009).

Para el entrenamiento de los datos y la información de las hojas, cada árbol  $T$  se construye sobre la base de un conjunto de parches  $P_i = \{(I_i, c_i, d_i)\}$ , donde  $I_i$  es la apariencia del parche,  $c_i$  es la etiqueta de la clase del parche, y  $d_i$  el vector de desplazamiento.

Los parches de formación son la muestra de la colección de entrenamiento de las imágenes. Los parches incluidos en la muestra del fondo se asignan con el valor de la etiqueta de la clase  $c_i = 0$ , mientras que a los parches incluidos en la muestra del interior del objeto se asigna  $c_i = 1$ . A cada parche objeto se le asigna también un vector 2D de desplazamiento igual a la posición desde el centro de gravedad del recuadro de delimitación, al centro del parche.

Durante el entrenamiento, a todo nodo no hoja de cada árbol se le asigna una prueba binaria aplicable a la aparición de un parche  $T_i$ . Tanto para el entrenamiento como para las pruebas, los parches tienen un tamaño fijo, por ejemplo, 16x16 píxeles, y la aparición se define por la extracción de las características de los canales, que pueden corresponder a la intensidad, las respuestas de los filtros al aplicar las derivadas de desplazamiento, entre otras características. Por lo tanto, la aparición del parche se puede escribir como  $T_i = (I_i^1, I_i^2, \dots, I_i^C)$ , donde cada  $I_i^j$  es una imagen 16x16 y  $C$  es el número de canales.

Las pruebas del  $t$  en el parche  $t(T) \rightarrow (0, 1)$  se pueden definir de varias maneras, pero (Lempitsky, 2009) se decidió por simples ensayos binarios basados

en píxeles. Este ensayo es el definido por un canal  $a \in (1, 2, \dots, C)$ , dos posiciones  $(p, q)$  y  $(r, s)$  en la imagen de  $16 \times 16$ , y una desventaja  $\tau$  del valor real. La prueba  $t_{a,p,q,r,s,\tau}(T)$  se define como:

$$t_{a,p,q,r,s,\tau}(T) = \begin{cases} 0, & \text{if } I^a(p, q) < I^a(r, s) + \tau \\ 1, & \text{Otherwise.} \end{cases}$$

De manera general la construcción de los árboles del bosque de Hough sigue el framework común de bosques aleatorios (Breiman., 2001). El nodo construido se declara una hoja y la información de la votación de la hoja  $(C_L, D_L)$  se acumula y almacena. De lo contrario, un nodo no hoja se crea y una prueba óptima es elegida entre un grupo grande de las pruebas binarias generadas. El conjunto de parches de entrenamiento que ha llegado al nodo después, se divide de acuerdo a la prueba elegida en dos subconjuntos que se pasan a los dos hijos nuevos que se han creado, después de que la recursividad proceda.

El momento clave de los bosques de Hough es la forma y calidad en que se evalúa la prueba binaria. En este caso se seleccionaron las incertidumbres tanto en las etiquetas de las clases como en el desplazamiento de disminución de los vectores hacia las hojas. Para lograr esta reducción, se definen dos medidas de la incertidumbre para un conjunto de parches, la entropía y la compensación de la incertidumbre.

Después del entrenamiento, el conjunto de árboles para clases específicas se puede utilizar para detectar los recuadros de delimitación de las instancias de una clase en una prueba de una imagen mediante la transformada de Hough. En primer lugar, se supone que el tamaño del objeto y dimensiones de las cajas se fijan durante entrenamiento y pruebas.

Considerando un parche  $P(y) = (I(y), c(y), d(y))$  centrado a la posición  $y$  en una imagen de prueba, donde  $I(y)$  es el aspecto observado de la revisión,  $c(y)$  es la etiqueta de clase, y  $d(y)$  es el vector oculto de desplazamiento desde el cuadro delimitador del centro del objeto a  $y$  (significativo sólo en el caso  $c(y) = 1$  por el cálculo logarítmico de  $c(y) = 0$ ). Por otra parte, se tiene a  $E(x)$  que denota el

suceso aleatorio correspondiente a la existencia del objeto centrado a la posición  $x$  en la imagen (Lempitsky, 2009).

Ahora se tiene que almacenar la evidencia probabilística  $p(E(x)|I(y))$  donde la aparición  $I(y)$  del parche da lugar a la disponibilidad de  $E(x)$  en diferentes posiciones  $x$  en la imagen. Luego se distingue entre los dos casos: y si pertenece al cuadro delimitador de  $B(x)$  centrado en  $x$  o no. Si  $y \notin B(x)$ , se asume que  $I(y)$  no es informativo sobre  $E(x)$ , poniendo  $p(E(x)|I(y)) = p(E(x))$ . El procedimiento de detección simplemente calcula la imagen de Hough  $V$  y devuelve el conjunto de sus ubicaciones y los valores máximos  $(x, V(x))$  como la hipótesis de detección. Los valores  $V(x)$  sirven como medida de confianza para cada hipótesis.

## Resultados de investigaciones asociadas

Las ventajas que ofrece el método multinivel de segmentación combinada mostrado en el epígrafe 1.2.1 quedan de manifiesto aplicando una detección de bordes mediante el método de Canny (Valverde, 2006), permite resaltar la eficacia del método cuando se trata de comprobar cuáles son las regiones detectadas previamente. Así, en la Figura 7, se puede comprobar que los contornos de los objetos segmentados por el método combinado son más precisos.

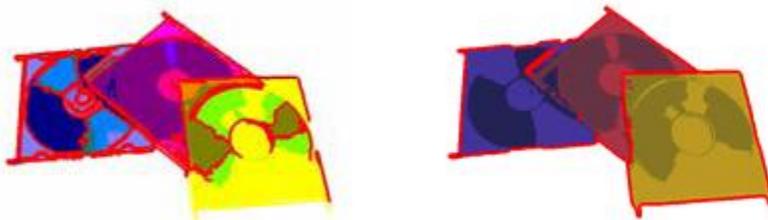
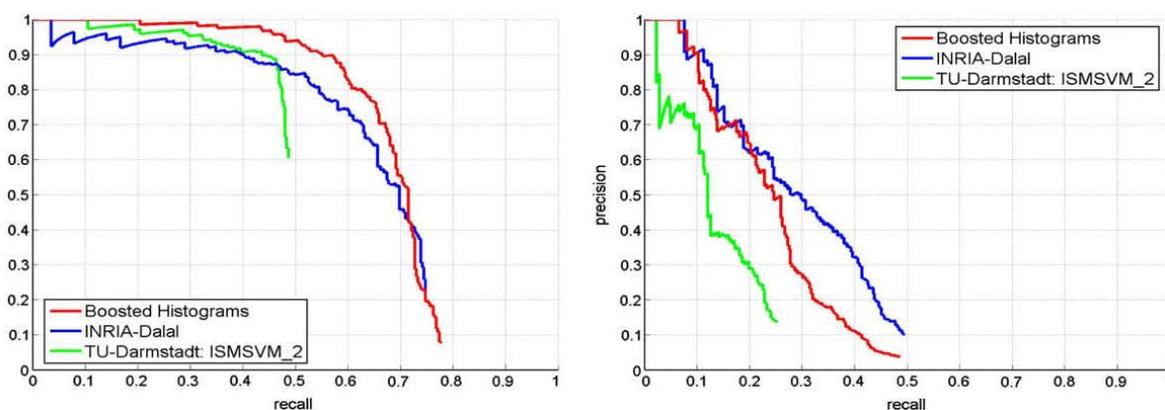


Figura 7. a) Detección de contornos de imagen segmentada RGB, b) Detección de contornos de imagen segmentada por combinación RGB-HSV.

El método expuesto en el epígrafe 1.2.2 difiere de la propuesta por Levi y Weiss (K. Levi, 2004) en tres aspectos principales: (i) se introduce un alumno débil por vectores característicos valorados, (ii) Se utiliza la posición de histograma en dependientes características y (iii) que artificialmente se amplía el entrenamiento positivo establecido. Para evaluar estas extensiones se comparó con algoritmos

de las competiciones (L. Van Gool, 2006) (Everingham, 2007) en la tarea de detectar objetos seleccionados en diferentes categorías: motos, bicicletas, personas, carros, caballos y vacas. El entrenamiento y los equipos de prueba contienen una variación sustancial de los objetos de las Bases de datos en términos de escala, constituyen la oclusión y la variabilidad dentro de la clase. (i) INRIA-Dalal/INRIA\_Douze basado en las características de HOG y SVM lineal (Triggs, 2005), (ii) TU-Darmstadt basado en los puntos de interés, ISM y SVM (M. Fritz, 2005.), entre otros. Las gráficas mostradas en la Figura 8 demuestran la eficiencia del método explicado en el epígrafe (Laptev, 2009).

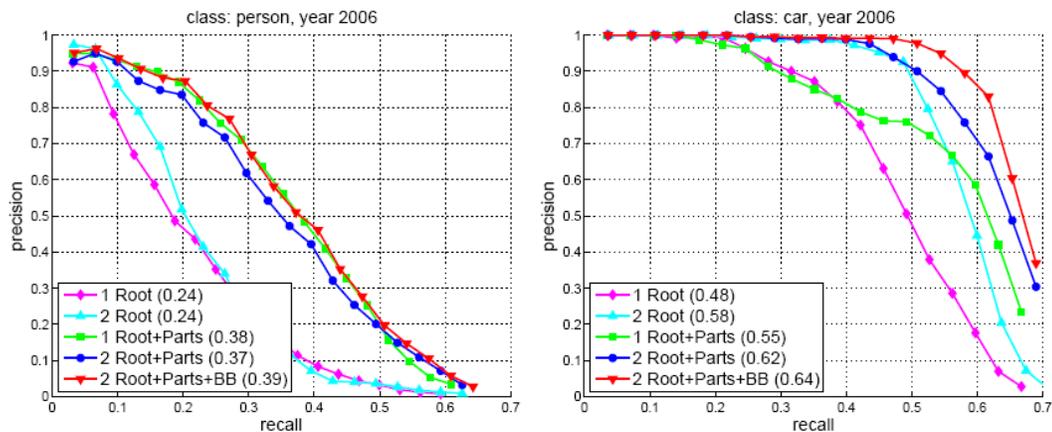


**Figura 8. Curvas Precision- recall para dos tareas de detección de objetos en la competición PASCAL VOC 2005 Challenge (derecha carro, izquierda motos). El método propuesto (boosted histogramas) es comparado con los mejores métodos reportados en (Everingham, 2007)**

Aplicando el método mostrado en el epígrafe 1.2.3 se tuvo como resultado, que las imágenes utilizadas para el entrenamiento deben de ser diferentes y bien definidas. A mayor cantidad de capas ocultas en la red, el resultado del aprendizaje será mejor para el entrenamiento de ellas. El umbral, la nitidez y posición de la cámara son factores influyentes en la identificación de los objetos a través de una red neuronal. Se pudo concluir que la red es limitada ya que solo identifica cuatro objetos, no pudiendo predecir que no es ninguno de ellos. Una forma de mejora y prueba de la validez de este algoritmo sería trabajar con videos en tiempo real, y no con imágenes pre-elaboradas.

La técnica del epígrafe 1.2.4 fue evaluada en diferentes aspectos con la base de datos de la competición Pascal 2006 (L. Van Gool, 2006) como se mencionaba

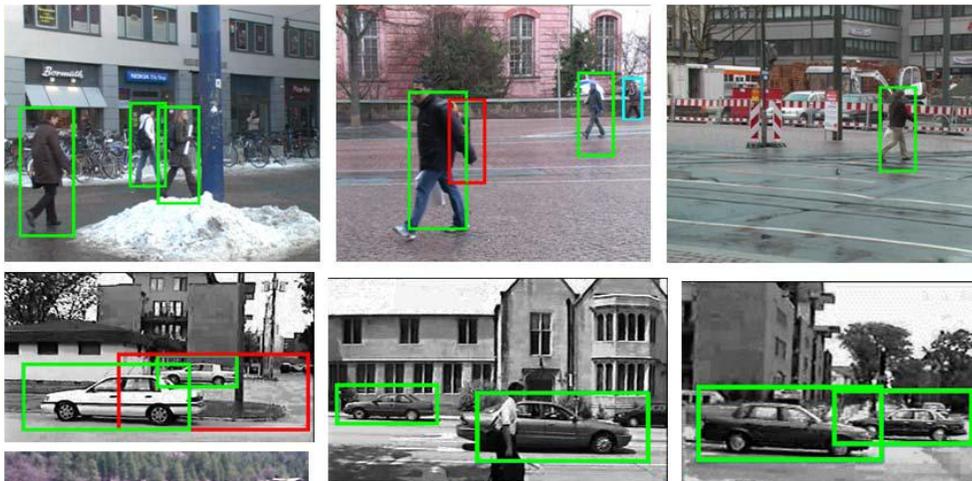
anteriormente. La Figura 9. Curvas Precision / Recall para modelos entrenados con las categorías carros y personas de la base de datos de la competencia PASCAL 2006 resume los resultados de diferentes modelos con las categorías persona y carros. Se entrenaron los modelos con dos componentes, primero simples y luego basados en partes deformables y luego además, se muestra el resultado del segundo componente con partes más las cajas de predicción y bordes (Pedro Felzenszwalb, 2009). Se puede apreciar una mejora significativa en la precisión de la detección comparada con los otros modelos y la bibliografía consultada, sin embargo se evidencia que los modelos compuestos muestran muy buenos resultados para la categoría de los carros, no siendo así para la de personas.



**Figura 9. Curvas Precision / Recall para modelos entrenados con las categorías carros y personas de la base de datos de la competencia PASCAL 2006**

El sistema presentado por Felzenszwalb y autores se basa en gran medida en los nuevos métodos para el entrenamiento discriminativo de clasificadores que hacen uso de la información latente. También utiliza métodos eficientes para hacer coincidir los modelos deformables a las imágenes. El sistema resultante es altamente eficiente y preciso según los resultados presentados, son ya capaces de representar a las clases de objetos muy variables, pero pretenden avanzar hacia modelos más robustos. El marco descrito permite la exploración de las estructuras latentes adicionales. Por ejemplo, se pueden considerar partes jerárquicas más profundas o modelos de mezcla con muchos componentes.

El algoritmo presentado en el epígrafe 1.2.5 fue evaluado en algunos conjuntos de datos que se muestran en la Figura 10. Resultados de la detección en las Bases de Datos TUD Peatones y UIUC-Escala, donde se ofrece una comparación de rendimiento con los métodos de detección relacionados, así como con los mejores resultados publicados previamente. Las curvas de rendimiento se han generado al cambiar la aceptación del umbral de la hipótesis  $V$  con una votación ( $x$ ).



**Figura 10. Resultados de la detección en las Bases de Datos TUD Peatones y UIUC-Escala**

Durante el entrenamiento los ejemplos positivos se reajustaron sobre la base de la misma altura, la mayor dimensión del recuadro delimitador fue de 100 x 100 píxeles por término medio durante un conjunto de datos, 20 000 pruebas binarias aleatorias fueron consideradas para cada nodo. Cada árbol fue entrenado con 25 000 parches positivos y negativos. Durante los primeros 5 árboles, los parches fueron incluidos en la muestra con probabilidades uniformes de todos los ejemplos disponibles positivos y negativos.

Para las tres bases de datos probadas en el artículo se utilizaron los mismos canales, 3 canales de color del espacio de color Lab, los valores absolutos de las derivadas de primer y segundo orden, y los nueve canales “HOG-like” (Triggs, 2005).

Para cualquier imagen del conjunto de datos evaluados, el algoritmo requiere, en las pruebas realizadas, un aproximado de 6 segundos para cada una de las cuatro operaciones: extracción de características, pasando por los parches de los árboles, acumulación de los votos, y por último procesamiento de las imágenes Hough. (720x576 de resolución; 4 escalas y una moderna CPU)

Anteriores enfoques han utilizado los bosques aleatorios como códigos discriminativos (F. Schroff, 2008) (J. Shotton, 2008) . Hough se diferencia del resto en cuanto al modo de almacenar los desplazamientos de los parches y su aprovechamiento para el tiempo de ejecución en la votación.

Estos bosques pueden ser eficientemente utilizados para detectar instancias de clases en las imágenes naturales, con la seguridad que no sólo es superior a los métodos anteriores relacionados con el estado de la técnica para varios conjuntos de datos, ver anexo 2. Aparte de la precisión, el uso de bosques aleatorios potencialmente permite una implementación muy eficiente en tiempo.

Esta versión sin optimizar, toma varios segundos por imagen, la aceleración de los factores registrados en la aplicación GPU de los bosques aleatorios en (Sharp, 2008) sugiere que casi en tiempo real el rendimiento es alcanzable, por lo que pudiera ser un elemento a tener en cuenta. Algunas sugerencias también serían la creación de bosques de Hough multi-clases, así como probar con los conjuntos de datos de las competiciones de PASCAL COV.

## **Conclusiones parciales**

Después de analizar algunas técnicas de resultados internacionales utilizadas en la detección de objetos, se puede resumir que tanto las técnicas de modelación basadas en segmentación e histogramas, como las que utilizan componentes de soporte de inteligencia artificial, se caracterizan por su efectividad y eficiencia en los resultados que ofrecen, sobre todo condicionados en pruebas preparadas para el algoritmo en cuestión.

Las propuestas que combinan el uso de varios criterios han permitido jerarquizar buenos resultados en las altas competiciones PASCAL VOC en sus diferentes ediciones, mejorando en cada año los tiempos de respuesta y la calidad de las selecciones resultantes del proceso de detección. Un inconveniente encontrado en la bibliografía consultada es lo relacionado con las aplicaciones de las técnicas, siempre basadas en ambientes condicionados y datos en las mismas estructuras (competencias), no siendo fiables con aplicaciones reales.

La tabla que se muestra a continuación relaciona los trabajos que han sido premiados en competiciones a nivel mundial sobre detección de objetos en videos.

Autores	Trabajo
Visual Object Classes Challenge 2009 (VOC2009)	
Pedro Felzenszwalb, Ross Girshick, David McAllester ( <i>University of Chicago; Toyota Technological Institute at Chicago</i> )	Discriminatively Trained Deformable Part Models
Andrea Vedaldi, Varun Gulshan, Manik Varma, Andrew Zisserman ( <i>University of Oxford; Microsoft Research India</i> )	Multiple Kernels for Object Detection
Koen van de Sande, Theo Gevers, Cees Snoek. ( <i>University of Amsterdam, Amsterdam, Netherlands</i> )	Color Descriptors
The 10th European Conference on Computer Vision (2008)	
Michael Isard and Andrew Blake ( <i>Department of Engineering Science, University of Oxford</i> )	Contour Tracking by Stochastic Propagation of Conditional Density
Matthew B. Blaschko and Christoph H. Lampert ( <i>Max Planck Institute for Biological Cybernetics, Tübingen, Germany</i> )	Learning to Localize Objects with Structured Output Regression.

**Tabla 1. Trabajos premiados en las últimas competiciones en el campo de la Visión por Computador en algunos eventos y competencias de gran prestigio**

## **CAPÍTULO 2**

### **Diseño de la metodología**

La representación conceptual y gráfica de sistemas o procesos constituye un elemento esencial e inseparable de toda actividad científica. En este trabajo, el término modelación se refiere al proceso de representación de la interrelación de los componentes de la metodología que se propone.

En el presente capítulo se explica el proceso de construcción de los procesos y modelos propuestos. En primer lugar, se propone un flujo escenificado en el diagrama del primer epígrafe, cuyo objetivo fundamental es el de maximizar la fiabilidad en los resultados de comparación de los objetos en secuencias de video y minimizar el tiempo de ejecución del procedimiento. Posteriormente, se explican las técnicas propuestas en cada uno de los componentes dentro del proceso de detección.

### **Estructura de la Metodología**

La Figura 11 describe los flujos de análisis propuestos, a los cuales se someten las secuencias de video para extraer las características principales, con el fin de detectar objetos específicos u objetos de manera general para caracterizar videos.

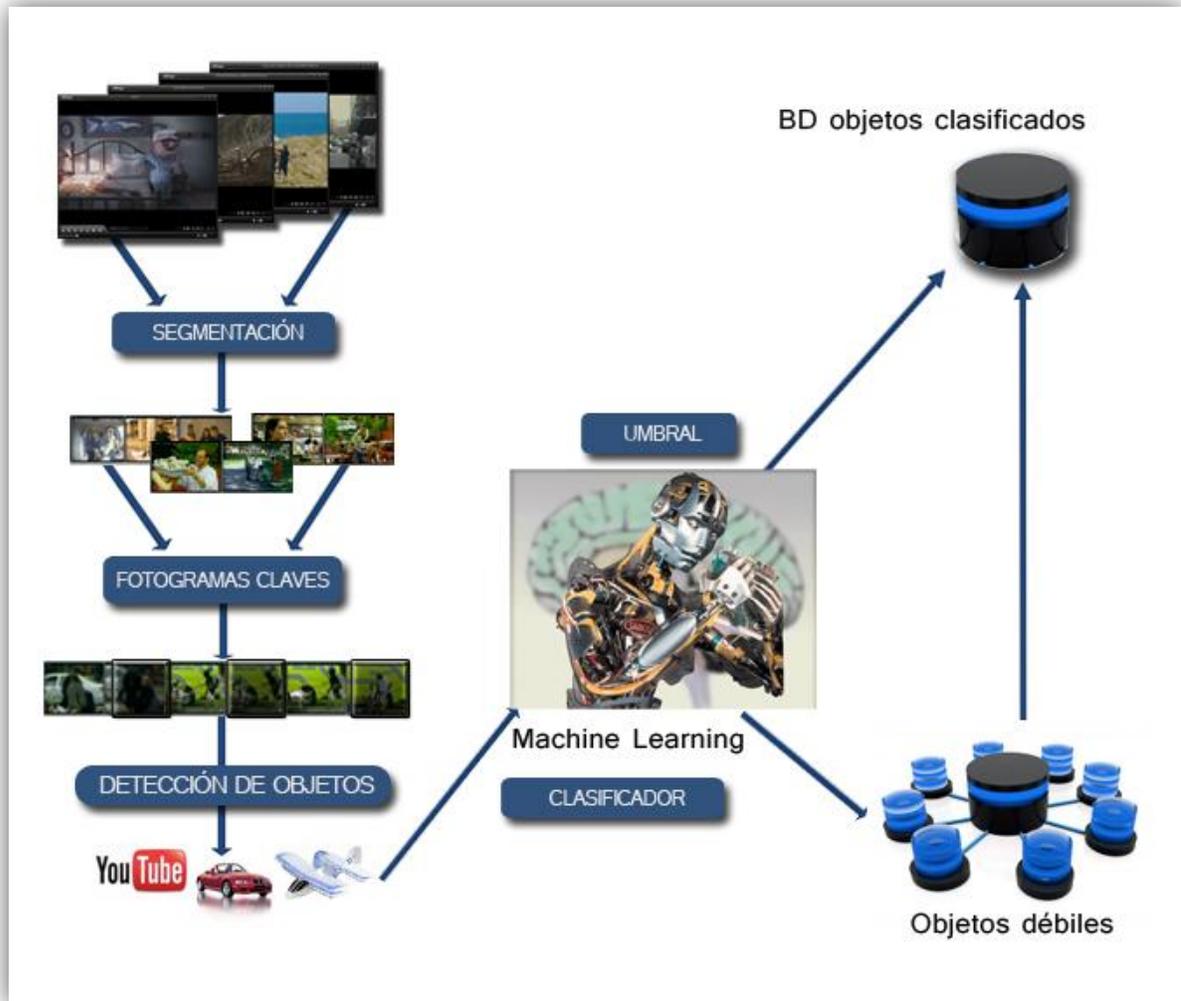


Figura 11. Diagrama de la Metodología propuesta

## Descripción de los componentes de la metodología

### 2.2.1. Segmentación de Videos

La gran mayoría de los sistemas de procesamiento de imágenes en el mundo, parten de diversas técnicas de segmentación por las facilidades descritas anteriormente en las definiciones. La metodología propone dividir los videos a partir de segmentaciones basadas en tomas o secuencias para ganar en eficiencia y rapidez, contando con diferentes hilos de procesamiento, determinados por cada

uno de los videos obtenidos, sin necesidad de hacer todo el resto del procesamiento al video completo.

#### *2.2.1.1. Detección de cambios de toma con información de contenido visual y auditivo.*

Una de las formas más comunes de segmentar contenido audiovisual consiste en la detección de cambios de toma. Esta tarea ha sido enfocada tradicionalmente desde el punto de vista de la señal de video, aunque hay algunos autores como (Chen, 2002) (Zhou, 2003), que emplearon una combinación de la información obtenida del audio y la imagen de forma que el sistema contara con una mayor robustez, obteniendo buenos resultados al implementar esta innovadora solución.

El algoritmo seleccionado para la división de los videos a analizar ha sido la detección de cambios de toma en contenidos audiovisuales desde dos perspectivas complementarias, por un lado la información existente en la parte visual y por el otro la información de audio. La extracción de características de la parte visual se realiza a través de los descriptores GoF/GoP y el color escalable MPEG-7, mientras que para la extracción de la información de la parte de audio se emplea BIC (Criterio de información Bayesiana de sus siglas en inglés) (Alejandro Abejon, 2009).

La técnica implementada realiza una detección de tomas mediante un modelo general, es decir, se busca que dicho modelo detecte tanto los cambios de toma abruptos como los graduales. Este sistema es equivalente a realizar una detección de cambio de toma por histograma de color, ya que los coeficientes sobre los que se aplican las medidas de dispersión miden la distribución de color en cada fotograma analizado.

Las medidas de dispersión evaluarán el grado de similitud entre fotogramas, dichas medidas se aplican sobre los coeficientes resultantes del descriptor. En lugar de calcular los coeficientes para todos los fotogramas se realiza para una de cada 25 de ellos, es decir un fotograma por segundo. Con este procedimiento se consiguen dos objetivos, por un lado reducir la carga computacional y por otro la

detección de cambios graduales. Como medidas de dispersión se emplearon 3 de las que mejor comportamiento presentaban en (Bescós, 2004).

El proceso seguido para obtener la información de audio consta de dos pasos básicamente:

- Separar el contenido de audio del video. Se realiza una parametrización MFCC<sup>9</sup> (Mel Frequency Cepstral Coefficients) de 13 coeficientes.
- Aplicar el algoritmo de detección de cambios, BIC (Gopalakrishnam, 1998) (Criterio de información Bayesiana, de sus siglas en inglés): Esta técnica busca puntos de cambio dentro de una ventana, para ello se comprueba si los datos de la ventana se modelan mejor con una única distribución (no hay punto de cambio) o con dos distribuciones (punto de cambio). Si se encuentra un punto de cambio se reinicia la ventana y se vuelve a iniciar la búsqueda a partir de ese punto. Si por el contrario no se encuentra punto de cambio se incrementa la ventana y se vuelve a realizar la búsqueda.

Antes de llevar a cabo la combinación de la información obtenida del audio y del video se realiza un filtrado de los resultados para subsanar los posibles errores en la detección. Este filtrado consiste en eliminar cambios de toma sucesivos que tengan entre sí menos de un segundo de duración. Se presupone que un cambio de toma por muy rápido que sea debe durar más de un segundo, por lo que los cambios de este tipo que resulten del algoritmo serán eliminados (Alejandro Abejon, 2009).

Una vez realizado este filtrado se pasa a combinar la información de los dos métodos para obtener un resultado global más robusto. Se puede tener en cuenta dos combinaciones: la más restrictiva de ellas (combinación &), consiste en indicar sólo aquellos cambios de toma en los que la información obtenida del audio y del video coincida. Sin embargo la conjunción es demasiada restrictiva por lo que solo se utilizará la (combinación +), que no es más que la agrupación de los cambios

---

<sup>9</sup> Coeficientes de la representación del espectro a corto plazo de la intensidad de un sonido, basado en una transformación lineal del coseno de un espectro de potencia de registro en una escala lineal de la frecuencia de Mel.

indicados por cada uno de los métodos por separado. Una vez hecho esto se realizará el proceso de filtrado mencionado anteriormente, pero con menos de tres segundos de duración.

### **2.2.2. Extracción de fotogramas claves**

Sería el primer paso a aplicar a cada uno de los videos segmentados del video original en forma paralela a través de hilos diferentes de procesamiento. Un video cuenta con un determinado frame rate<sup>10</sup>, en dependencia de ello tiene entre 23 ó 30 frames x segundo, lo que quiere decir que 5 minutos de videos solamente tendrían como promedio 7500 frames. Se propone entonces no analizar todos los frames para encontrar los objetos que se estén buscando o los que se quieran caracterizar, sino extraer los fotogramas claves (keyframes) con que cuente el video.

Este procedimiento tiene un inconveniente y es que se puedan quedar frames con información importante (objetos buscados) entre dos keyframes, sin embargo este porcentaje de omisión puede ser minimizado por la selección adecuada del umbral. Se realizaría una prueba con extensas cantidad de videos para obtener este umbral planteado que garantice eficientes resultados.

#### *2.2.2.1. Algoritmo para la extracción de fotogramas claves en sumarización de videos*

La realización de resúmenes de video, destinados a reducir la cantidad de datos que deben ser examinados con el fin de recuperar la información deseada de la información en un video, es una tarea esencial en el análisis de video y aplicaciones de indexación, en el caso de este trabajo para reducir el costo computacional a la hora de detectar objetos. Para esta tarea dentro de la metodología se propone un enfoque innovador para la selección de los fotogramas claves representativos de una secuencia de video (G. Ciocca, 2006). Al analizar las diferencias entre dos imágenes consecutivas de una secuencia de video, el

---

<sup>10</sup> Velocidad de fotogramas con que ha sido transcodificado un video

algoritmo determina la complejidad de la secuencia en términos de cambios en el contenido visual expresadas por descriptores de frames diferentes.

Los fotogramas claves se extraen mediante la detección de puntos de curvatura en la curva de las diferencias de frames acumuladas. Otra ventaja es que puede extraer los fotogramas clave sobre la marcha: los puntos de curvatura se pueden determinar cuando el cálculo de las diferencias de los frames y los fotogramas claves pueden extraerse tan pronto como el segundo punto de alta curvatura ha sido detectado.

Antes de describir el algoritmo seleccionado se mencionarán algunas técnicas para extraer fotogramas claves en el mundo, primeramente se tiene la selección del primer fotograma en un shot como el fotograma clave del fragmento del video (Tonomura Y., 1995), (Rui Y., 1999) utiliza el primero y el último. Otros seleccionan intervalos fijos dentro de una secuencia de video para obtener los fotogramas claves (Pentland A., 1995). Sin embargo ninguna de estas técnicas tiene en cuenta la dinámica de un contenido visual que tiene un shot, sino que se basan en la información sobre la secuencia dada por límites.

El umbral de selección o la medida de fidelidad en el algoritmo está dado por la comparación en cada fotograma clave del resumen con el resto de los fotogramas en una secuencia de video, es definida por la distancia semi-Hausdorff<sup>11</sup>. En general, un descriptor visual único no puede captar todos los detalles pictóricos necesarios para estimar los cambios en el contenido visual de los fotogramas.

Al definir lo que es una buena representación pictórica de un fotograma, hay que tener en cuenta las propiedades de color, textura y estructurales. Para superar este problema de la representación del fotograma presente en varios algoritmos que realizan comparación entre frames, en este algoritmo (G. Ciocca, 2006), se calculan tres descriptores diferentes: un histograma de color, un histograma de la dirección del borde y por último las estadísticas wavelet. Estas características han

---

<sup>11</sup> Mide hasta qué punto dos subconjuntos están alejados uno del otro en un espacio métrico.

sido seleccionadas para tres propiedades básicas: la similitud de la percepción, eficacia (las características pueden ser rápidamente computarizada), y la economía.

El algoritmo de selección de fotogramas claves que se propone selecciona dinámicamente los frames representativos mediante el análisis de la complejidad de los eventos descritos en un shot, en términos de cambios pictóricos. La diferencia de los valores obtenidos en los frames inicialmente se utiliza para construir una curva de las diferencias acumuladas que describe cómo cambia todo el contenido visual del shot mediante el análisis de los fotogramas, una indicación de la complejidad de la secuencia: las pendientes fuertes indican cambios significativos en el contenido visual, debido a un objeto en movimiento, el movimiento de la cámara, o el registro de un gran evento dinámico. Los fotogramas claves se identifican en la curva de las diferencias acumuladas de los frames como aquellos puntos de la curva donde los ángulos son más agudos. Para detectar los puntos de curvatura se propone utilizar el algoritmo planteado por (Z, 1999).

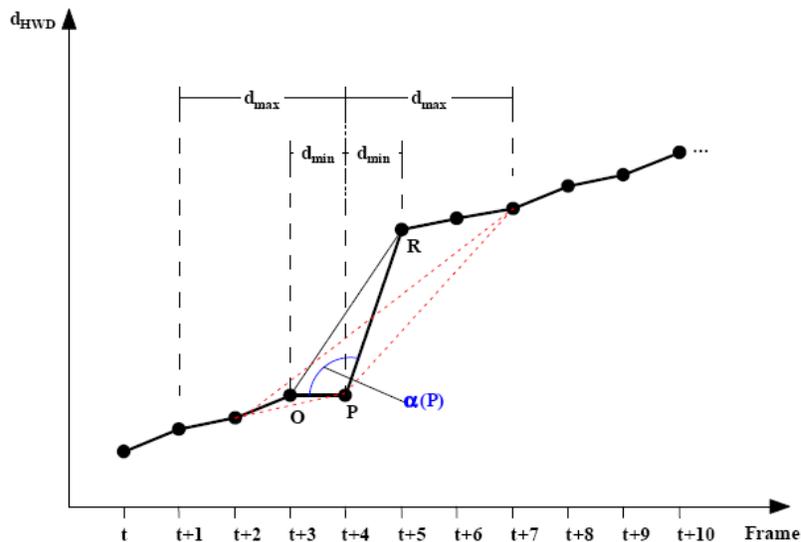


Figura 12. Algoritmo de detección de los puntos de alta curvatura

En la aplicación se han definido los puntos mínimos  $d_{min}$  distancia siempre igual a 1 y, en consecuencia los dos únicos parámetros que influyen en los

resultados del algoritmo son  $dmáx$  y  $Amax$ . El parámetro más importante es  $Amax$  que controla el conjunto de ángulos admisibles: un alto valor de  $Amax$  dará lugar a mayor cantidad de puntos incluidos en el conjunto de candidatos de alta curvatura de puntos, mientras que un valor más bajo indica que sólo ángulos muy agudos deben ser considerados.

### **2.2.3. Detección de Objetos**

En este flujo de la metodología, se aplicaría la técnica para la detección de objetos seleccionada al resumen del video realizado con los fotogramas claves. Al aplicar esta técnica en una catalogación automática de un video para indexar contenido a una base de datos, sería solamente en este componente del algoritmo, teniendo en cuenta que sólo saldría la información indexada en la BD de la aplicación, pero si lo que se quiere es buscar un objeto determinado, habría que aplicarla a priori para guardar en la base de datos de objetos clasificados en caso de que no existan, el elemento a buscar, para tener entonces un punto de comparación y con los mismo rasgos básicos del objeto en un keyframe determinado.

#### *2.2.3.1. Detección de objetos a partir de modelos multi escalas de partes deformables.*

Después de analizadas las técnicas expuestas en el capítulo 1 durante la fundamentación teórica, se seleccionó el algoritmo basado en los bosques de imágenes de Hough con árboles binarios de decisión que utiliza en una de sus variantes (Lempitsky, 2009), en el epígrafe 1.2.5. Esta técnica es muy rica en datos de recuperación que clasifican los frames en una secuencia de video, ideal para la metodología propuesta en este trabajo e imprescindible para el auto aprendizaje de las aplicaciones que lo utilicen.

Como se mencionó anteriormente esta técnica utiliza métodos eficientes para la comparación de imágenes con clases de objetos multiescalas a partir de árboles binarios de búsqueda con una gran cantidad de datos a tener en cuenta. Se puede considerar por sus resultados, eficiente y preciso; es posible que represente a

clases de objetos muy variables, pero se pretende avanzar según la literatura hacia modelos más robustos, por lo que en la práctica se pueden mejorar los resultados a obtener durante la implementación.

Se pueden considerar además partes jerárquicas más profundas o modelos mezclados con varios componentes para la metodología propuesta creando un valor adicional que explote mejor los resultados de estas técnicas. No obstante se consideran los bosques de imágenes de Hough, una variante bastante novedosa en el campo de la visión por computador, que mejorando su tiempo de ejecución y la selección de los canales y elementos de los parches durante la comparación, se pueden obtener resultados muy interesantes que se verán en el capítulo 3.

#### **2.2.4. Machine learning**

La máquina de aprendizaje se propone con el objetivo de que sea la que contenga el discriminador de la algoritmia del flujo. Tendría dos componentes esenciales, el clasificador para determinar si el objeto que se está comparando, que se obtuvo obtenido de un fotograma clave, es alguno de los guardados y el segundo, el de aprender a discriminar entre un grupo de objetos clasificados como débiles para que en algún momento puedan agruparse y conformar un objeto que pase a estar clasificado según los descriptores que se contengan del objeto analizado y los contenidos en la base de datos de objetos débiles.

En este componente además del umbral necesario a seleccionar, se pueden tener en cuenta para la toma de decisiones técnicas o algoritmos inteligentes artificiales (como son una máquina de soporte vectorial o una red neuronal). La etiquetación de la acumulación de objetos débiles partiría de una alarma en los sistemas que lo implementan, ya que se guardarían las características discriminativas del posible objeto que pasa hacia la base de datos comparativa o principal.

#### **2.2.5. Base de datos**

El componente base de datos en la metodología es imprescindible para almacenar los objetos clasificados con la técnica para la detección de objetos, los

cuales son la base de comparación cuando se analiza un video. Además, habría otra base de datos para tener los objetos débiles (no llegan a coincidir con ningún objeto clasificado, pero un sistema implementado a partir de la metodología entiende que tiene características potenciales para serlo). La máquina de aprendizaje, a partir de dos momentos de umbralización, cantidad de objetos similares y el grado de similitud traslada hacia la base de datos de objetos clasificados estos débiles.

## **Conclusiones Parciales**

Durante la creación de una técnica de reconocimiento y detección de objetos, se debe prestar especial interés al proceso de selección de estructuras y descriptores para la representación y comparación de los mismos. En términos generales las técnicas deben caracterizarse, entre otras propiedades, por brindar un acceso eficiente a la información espacial contenida para realizar los análisis correspondientes. Generalmente los algoritmos de detección de objetos o de análisis de video tienen un costo  $O(\log n)$ , siendo  $n$  el total de estructuras contenidas que se necesitan analizar, lo que implica una especial atención al costo computacional del algoritmo y la estructura correspondiente. En este sentido se tratará de minimizar al máximo la implementación futura de la metodología basada en los errores que puedan presentar las técnicas analizadas.

Permitir la localización exacta de regiones contenedoras de la información que garantice la navegación eficiente sobre los videos analizados será punto también de atención. A diferencia de sus similares, en la metodología propuesta, la búsqueda de la información en los fotogramas más representativos de los videos y la separación en varios hilos de procesamiento, constituye una optimización importante en la búsqueda de los objetos.

Se debe tener en cuenta, que la selección inadecuada del umbral de la extracción de los frames con mayor cantidad de objetos en la secuencia de video, podría traer pérdida de información importante porque requiere de un mínimo de

cantidad de memoria para el procesamiento, que pudiera aprovecharse en la detección de objetos de manera general.

## **CAPÍTULO 3**

### **Análisis de los resultados**

Técnicas de reconocimiento y detección de objetos como las tratadas en la sección 1.2, han demostrado ser eficientes y flexibles en los resultados obtenidos en las competiciones aplicadas. En el presente capítulo se exponen un conjunto de resultados obtenidos a partir de pruebas realizadas a las técnicas que han sido propuestas en la metodología. Se mostrarán fotogramas con el objeto a detectar al aplicar el algoritmo propuesto y el resultado obtenido, demostrando la factibilidad de su uso en el tratamiento del problema planteado.

Además, en el presente capítulo, al no probar en su totalidad el uso e implementación de la metodología, se aplican algunas encuestas para obtener criterios de expertos que evalúan la posible aplicación de la misma.

### **Herramientas de ensayo**

Para realizar las pruebas prácticas con las técnicas propuestas se utilizaron el software MATLAB R2007a y la integración del entorno de desarrollo QT Creator con la librería OpenCV.

MATLAB es el nombre abreviado de “Matrix Laboratory”, es un programa para realizar cálculos numéricos con vectores y matrices (la mejor y más eficiente forma de trabajar las imágenes). Como caso particular se puede también trabajar con números escalares –tanto reales como complejos–, con cadenas de caracteres y con otras estructuras de información más complejas. Una de las capacidades más atractivas es la de realizar una amplia variedad de gráficos en dos y tres dimensiones. MATLAB tiene también un lenguaje de programación propio.

Dentro de la instalación del MatLab se cuenta con una serie de herramientas adicionales para el trabajo en diferentes campos y aplicaciones matemáticas con una gran variedad de funciones implementadas que optimizan mucho el trabajo para el cual se esté usando el software. El Toolbox para el procesamiento de imágenes, será utilizado en este trabajo, proporciona un amplio conjunto de

algoritmos estándar de referencia en el campo de la visión por computador y herramientas gráficas para el procesamiento de imágenes, análisis, visualización y desarrollo de algoritmos.

Puede restaurar imágenes ruidosas o degradadas, mejorar las imágenes de inteligibilidad mejorada, extraer características, analizar las formas y texturas, y registrar dos imágenes. La mayoría de las funciones están escritas en el lenguaje abierto MATLAB, que brinda la posibilidad de inspeccionar los algoritmos, modificar el código fuente y crear sus propias funciones personalizadas.

En cuanto a OpenCV proviene de los términos en inglés “Open Source Computer Vision Library”, es una librería de tratamiento de imágenes, destinada principalmente a aplicaciones de visión por computador en tiempo real. Desde que apareció su primera versión alfa en el mes de enero de 1999, se ha utilizado en infinidad de aplicaciones. Desde sistemas de seguridad con detección de movimiento, hasta aplicativos de control de procesos donde se requiere reconocimiento de objetos. Esto se debe a que su publicación se da bajo licencia BSD, que permite que sea usada libremente para propósitos comerciales y de investigación con las condiciones en ella expresadas ([Medrano, 2008](#)).

Open CV es multiplataforma, Existiendo versiones para Linux, Mac OS X y Windows. Esta librería, proporciona tipos de datos básicos para su utilización. También provee tipos de datos, funciones y algoritmos básicos de procesamiento introducidos como ayuda al programador, para hacer que el acceso a las informaciones de interés sea más simple.

Para probar la detección de objetos propuesta en la tesis, corazón de la metodología, se utilizó parte de un código ejemplo provisto por los autores del artículo "Bosques de Hough para la detección de objetos" el cual fue compilado para Linux y corregido en algunos aspectos de salida en el entorno de desarrollo QT Creator y lenguaje de programación C++. En el Anexo 3: Diagrama de clases de la técnica Bosques de Hough para la detección de objetos. se muestran las estructuras de las clases utilizadas, así como algunas de las imágenes de Hough

obtenidas al probar la aplicación con las bases de datos de Weizzman Horse y Airplanes Dates Trevick como se muestran en el Anexo 4: Aplicación de la técnica a dos imágenes de dos bases de datos de imágenes.

## **Experimentos**

Con el objetivo de mostrar la efectividad de la extracción de fotogramas claves, se utilizará el software MatLab con su Toolbox para el procesamiento de imágenes. Fue seleccionado un video arbitrario de dibujos animados que puede aparecer en las aplicaciones de catalogación automática.

Como se están realizando experimentos se utilizó el VirtualDub v1.9.8 para tomar las muestras de fotogramas de una secuencia del video antes mencionado y no tener que realizar la captura y extracción de fotogramas desde los ejemplos de códigos en Matlab, así como bases de datos de imágenes variadas para la detección de objetos.

### ***3.2.1. Extracción de fotogramas claves***

Para probar la propuesta de la extracción de fotogramas claves se utilizó una secuencia de 350 fotogramas, se realizó una pequeña prueba del algoritmo con Matlab y los resultados fueron buenos teniendo en cuenta que lo que se utilizó fue un dibujo animado, que tiene ventajas sobre los videos de imágenes reales en cuanto a ausencia de sombras forzadas y una alta ganancia en los colores y bordes, ya que el análisis se basa en la complejidad de los eventos descritos en la secuencia en términos de cambios pictóricos.

De los 350 fotogramas se obtuvieron 13 fotogramas claves, de estos no hubo ningún falso negativo y solo hubo un falso positivo.

### ***3.2.2. Detección de los objetos***

Para evaluar la fiabilidad del algoritmo se aplicó el entrenamiento y prueba del mismo con el código brindado por los autores y una función en MatLab para la detección del objeto mediante un recuadro delimitador. Las bases de datos seleccionadas fueron las mismas expuestas por los autores, "Weizman Horse DB",

“TUD Pedestrian DB”, más una de motos de lado, “motorbike-testset TUD DB” y una de aviones “Airplanes Dates Trevick DB”. Estas últimas fueron entrenadas con 50 imágenes positivas y negativas respectivamente, parches de 16 x 16 píxeles y 200 ejemplos de estos para las imágenes negativas y positivas para la generación de los bosques de Hough. Para las dos primeras bases de datos se siguieron los mismos parámetros de entrenamiento y detección que proponían los autores para realizar la valoración de los resultados.

Para el entrenamiento se cargan las imágenes positivas y negativas y se siguen las siguientes funciones:

```
// Se inicializa el bosque con un número de árboles
CRForest crForest( ntrees );

// Se inicializa un generador aleatorio para las semillas
time_t P = time(NULL);
int seed = (int)P;

CvRNG cvRNG(seed);

// Se inicializa el dato para el entrenamiento
CRPatch Train(&cvRNG, p_width, p_height, 2);

// Se extraen los parches para el entrenamiento
extract_Patches(Train, &cvRNG);

// Entrenamiento del bosque con los parches
crForest.trainForest(20, 15, &cvRNG, Train, 2000);

// Se salva el bosque con los árboles generados
crForest.saveForest(treepath.c_str(), off_tree);
```

Con el bosque de entrenamiento salvado se aplica la detección con las imágenes de prueba que también tendrían tamaños definidos para los parches que deben coincidir con los de entrenamiento, para este caso 16 x 16 píxeles. Las funciones de detección serían las que siguen:

```
// Se carga el bosque para la comparación con cada imagen de prueba
crForest.loadForest(treepath.c_str());

// Se inicializa el Detector
CRForestDetector crDetect(&crForest, p_width, p_height);

// Se corre el detector
detect(crDetect);
```

Después de realizadas las pruebas, los resultados obtenidos evidencian una buena precisión como lo muestran las graficas siguientes, aunque se puede decir que para el caso de las personas, se comprobó que habría que utilizar otra variante de detectar el máximo en la imagen de Hough ya que cuando existe más de un objeto, sólo se mostraría el máximo en acumular votos de los que existen en la imagen original.

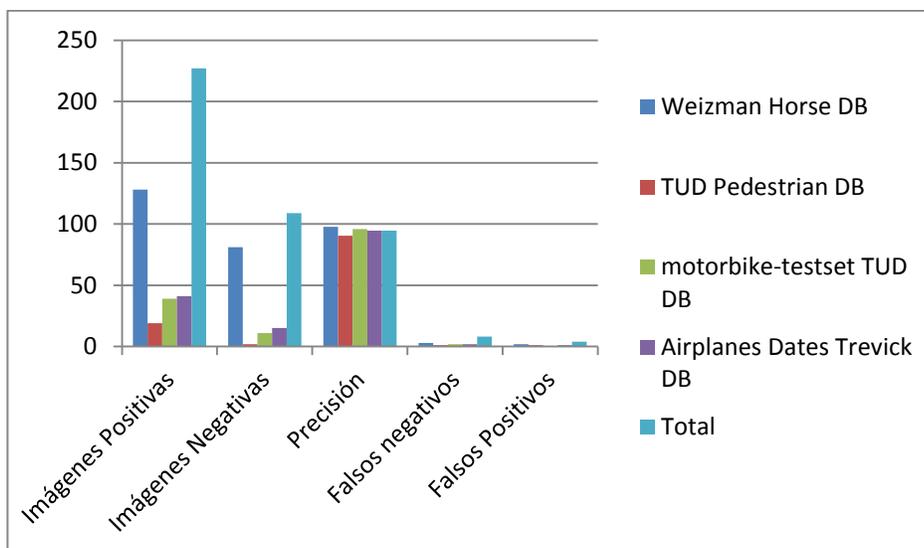


Figura 13. Gráfica con los resultados de los bosques de Hough al aplicarlo a 4 muestras de bases de datos de objetos.

Airplane DB								
Umbral	TP	FP	FN	TN	Total	Precisión	Recall	Acuraccy
65	150	12	1	22	185	0,93	0,99	0,93
70	149	7	2	27	185	0,96	0,99	0,95
75	145	2	6	32	185	0,99	0,96	0,96
80	137	0	14	34	185	1,00	0,91	0,92
85	133	0	18	34	185	1,00	0,88	0,90
Horse DB								
130	123	6	0	75	204	0,95	1,00	0,97

135	123	4	0	77	204	0,97	1,00	0,98
140	122	1	1	80	204	0,99	0,99	0,99
150	120	1	3	80	204	0,99	0,98	0,98
160	118	0	5	81	204	1,00	0,96	0,98
<b>MotorBike DB</b>								
80	98	7	0	15	120	0,93	1,00	0,94
90	97	6	1	16	120	0,94	0,99	0,94
100	97	2	1	20	120	0,98	0,99	0,98
110	96	2	2	20	120	0,98	0,98	0,97
120	95	0	4	21	120	1,00	0,96	0,97

Figura 14. Tabla con los valores de precisión y recal de 3 de las bases de datos probadas con el algoritmo.

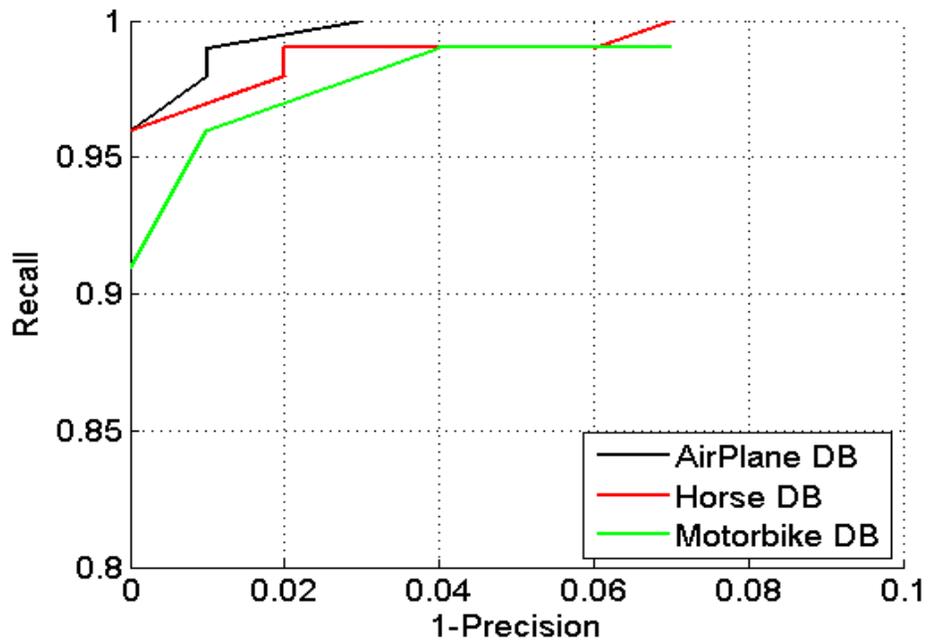


Figura 15. Grafica Precision – Recall de las muestras referenciadas en la tabla anterior

## Encuestas a expertos

Además de probar algunos códigos de la metodología, se utilizó el método Delphi, como procedimiento de estructuración de un proceso de comunicación grupal que es efectivo a la hora de permitir a un grupo de individuos, como un todo, tratar un problema complejo (Cabezas, 2004).

Este método consiste en la interrogación de un grupo de personas a las que se les llama expertos con la ayuda de cuestionarios, a fin de poner de manifiesto convergencias y opiniones. Las encuestas se llevan a cabo de manera anónima.

Suelen distinguirse tres etapas o fases fundamentales en la aplicación del método:

- Fase preliminar: Se delimita el contexto, los objetivos, el diseño, los elementos básicos del trabajo y la selección de los expertos.
- Fase exploratoria: Elaboración y aplicación de los cuestionarios.
- Fase final: Análisis estadísticos y presentación de la información.

El principal objetivo que se sigue con el método es que, basado en la experiencia de un grupo de expertos, los cuales trabajan en el procesamiento de imágenes en la Universidad de las Ciencias Informáticas (UCI) o poseen un buen nivel científico en otras áreas temáticas de la informática, se logre la validación de la propuesta de manera rápida con el fin de lograr la implementación de la metodología. Esto permitirá mejorar la calidad y los valores agregados de las plataformas que procesen videos en el departamento Señales Digitales del centro de Geoinformática y Señales Digitales (GEySED).

### ***3.3.1. Tipo del Método Delphi a utilizar***

Delphi se puede clasificar en varios tipos: Conducción (convencional y Computador), por objetivo (Proyección y Política) y de otros tipos como el de cara-cara (Cabezas, 2004). Para decidirse por uno de los métodos se tuvo en cuenta la rapidez de las respuestas y la disminución del tiempo.

Se decidió que el cuestionario a realizar fuera llevado personalmente a cada una de las personas integrantes del panel lo cual disminuye el tiempo y aumenta la rapidez de las respuestas debido a que el panelista puede consultar cualquier duda con respecto a la preguntas del cuestionario. Este método se le conoce como cara-cara (Cabezas, 2004) el cual tiene la ventaja de disminuir el porcentaje de defección del experto.

### **3.3.2. Selección de los expertos**

Para la selección de los expertos se tuvo en cuenta la efectividad en su trabajo profesional, como característica fundamental para la solución del problema tratado y por ser la que logra mayor objetividad en la evaluación del resultado. También se tomó en cuenta su disposición para realizar el cuestionario, lo que fue esencial para expresar la calidad de la validación.

Para aumentar la eficacia, normalmente se eligen expertos reconocidos por su prestigio profesional y con conocimientos acerca del tema. Por ser el tema de la investigación poco trabajado aún en la UCI, no todos los expertos seleccionados tienen un alto nivel de conocimientos en la temática pero si son compañeros de nivel científico y reconocimiento en el área de aplicación, no obstante no es obligatorio que respondan todas las preguntas presentadas en la encuesta, solamente con las que se sientan en criterio de opinar con base. El panel de expertos está compuesto de la siguiente manera:

- Un doctor del Departamento de Arquitectura por Computadoras de la Universidad de Málaga, que trabaja en el área de la investigación.
- Dos doctores de la facultad.
- Tres másteres de la facultad.
- Cuatro miembros de la línea de investigación de procesamiento de imágenes y video del departamento Señales digitales, unido a dos compañeros del centro de la facultad 10 que trabajan en investigaciones de temas afines.

### **3.3.3. Elaboración de la Encuesta**

El cuestionario se hizo con planteamientos de mediana longitud, debido a que si es un cuestionario muy largo puede inducir al desinterés de los expertos y uno muy corto puede provocar que no se cumplan los objetivos del mismo, además teniendo en cuenta que se realizaron algunas pruebas prácticas a la metodología. Durante la realización, se hizo énfasis en las preguntas de votación las cuáles consisten en poner más de una alternativa con el objetivo de que el experto pueda decidirse por una, aunque en algunas preguntas se incluyeron, para efectuar un chequeo a las coherencias de las personas que forman parte del panel, características de las preguntas de control que consisten en hacer la misma pregunta de forma diferente.

En el Anexo 5: Preguntas de la Encuesta a los Expertos. se muestra el cuestionario que cuenta con 6 enfocadas en darle cumplimiento a los siguientes criterios fundamentales:

- Establecimiento (pregunta 2)
- Impacto (pregunta 3)
- Generalidad (pregunta 4)
- Valor científico (pregunta 5)

Las preguntas 1 y 6 se hicieron con el objetivo de conocer la especialidad profesional del experto y acumular un grupo de sugerencias que podrían servir también al mejoramiento del desarrollo futuro de la metodología.

Los expertos que conformaron el panel recibieron un resumen de la tesis para emitir su criterio respecto al mismo y responder las preguntas. Los resultados se procesaron de forma manual pues la cantidad de información a manejar no es muy extensa ni complicada. Para llegar a un consenso con los expertos se utilizó la siguiente regla que es adaptada para Delphi.

Se entenderá por consenso, cuando una de las alternativas acumula al menos el 70% del total de las alternativas ponderadas por nivel de confianza y grado de los expertos en el tema.

- Pregunta 1

La pregunta fue respondida por 12 entrevistados que dentro de su función está la investigación como proceso fundamental en la facultad. Entre ellos se encuentran directivos de la facultad y el centro, e investigadores del departamento Señales Digitales del Centro GEySED.

- Pregunta 2

La pregunta la respondieron 11 entrevistados lo que representa el 91.6 % del panel. Como el 100 % de los panelistas concuerdan que es necesario el empleo de una metodología para la detección de objetos en los sistemas audiovisuales del departamento y así estandarizar el trabajo en la línea de investigación, se logra una concordancia entre los expertos.

- Pregunta 3

La pregunta fue respondida por 10 entrevistados lo que representa el 83.3 % del panel. Como el 90 % de los panelistas concordaron que las posibilidades de aplicación son buenas, se logra una concordancia entre los expertos.

- Pregunta 4

La pregunta fue respondida por 10 entrevistados, lo que representa el 83.3 % del panel. De ellos el 80 % concuerda que es posible la implementación de la propuesta con un nivel medio, se puede decir que se logra un consenso entre los expertos.

- Pregunta 5

La pregunta fue respondida por 11 entrevistados, lo que representa el 91.6 % del panel. Como el 100 % concuerdan que la calidad de la investigación y propuesta de la metodología para la detección de objetos es buena, se puede decir que se logra un consenso entre los expertos.

- Pregunta 6

La pregunta de sugerencia fue respondida por 3 entrevistados, los cuales sugirieron que:

- Antes de comenzar a implementar completamente la metodología propuesta, sería prudente hacer un software de prueba de todos los componentes propuestos, así como la utilización de diversas clases de objetos para aumentar la fiabilidad del algoritmo de detección de objetos.
- Buscar la fusión del algoritmo de detección de objeto con los Bosques de Hough con algunos otros algoritmos de detección, como por ejemplo el Viola and Jones ([P. Viola, 2001](#)), utilizando aspectos fundamentales de sus ventanas deslizantes y detección temporal de características
- Utilizar una máquina de soporte vectorial para la extracción de características y aumentar las muestras espaciales a comparar de las diversas clases de objetos.

## **Conclusiones Parciales**

Las limitaciones en la disponibilidad de memoria y velocidad de procesamiento que imponen los actuales sistemas, es un elemento a tener en cuenta en el diseño e implementación de técnicas de procesamiento de video, por esto es que una aplicación de catalogación automática que logre detectar objetos, debe caracterizarse por realizar un uso racional y eficiente de estos recursos. El uso de operaciones recursivas para la extracción de características en las imágenes y estructuras de datos adaptables para los cálculos y aprendizaje, incide directamente en la disminución de los requerimientos de memoria de las técnicas.

El orden de ejecución de los algoritmos implementados para estos fines, debe garantizar un adecuado aprovechamiento de las velocidades de cómputo de los sistemas, factor que contribuye positivamente a la eficiencia del proceso de reconstrucción de las imágenes detectadas a mostrar.

Al concluir este capítulo, se puede decir que se ha validado en gran medida la metodología propuesta, utilizando una fusión entre el método Delphi y ejecución de códigos ejemplos de la extracción de fotogramas y detección de objetos

multiescalas. Los entrevistados concordaron en su mayoría con las respuestas a las preguntas y la evaluación de los cuestionarios aportó resultados interesantes para el futuro de la implementación de la metodología.

## Conclusiones

- La representación de un video a partir de varias secuencias de imágenes divididas por shots, y la posterior extracción de fotogramas representativos a tener en cuenta, reduce considerablemente el tiempo de ejecución de los algoritmos para la detección y segmentación de videos.
- La separación en hilos de procesamiento ayuda a la disminución de la complejidad a ejecutar en la GPU para el entrenamiento y la posterior detección de los objetos.
- La detección de objetos a partir de los bosques de Hough ha demostrado tener algunos problemas con los elementos multiescalas ([Lempitsky, 2009](#)), sin embargo se ha probado que resulta altamente eficaz para clases específicas en el tratamiento del problema planteado, y más cuando se pueden corregir algunos inconvenientes con la catalogación semiautomática.
- Se han definido los conceptos que soportan el desarrollo de una metodología para detectar objetos en secuencias de video, se describieron las características y procesos con que se propone cuenta la misma. Se diseñó un diagrama que describe el funcionamiento de la metodología con el fin de facilitar su comprensión y posterior implementación.

## Recomendaciones

- Continuar el estudio de las diversas técnicas que existen en el mundo, para aplicar y proponer mejoras en las propuestas en la metodología, ya que la segmentación temporal de videos, extracción de fotogramas claves y detección de objetos, son problemas que en la actualidad no están completamente resueltos.
- Incrementar la eficiencia del algoritmo para que funcione para múltiples objetos en una imagen, además de versatilidad de escalas y niveles de características a tener en cuenta por cada objeto almacenado durante el entrenamiento.
- Con el análisis realizado (comparaciones mundiales, criterios de expertos e implementación de algunos componentes de la metodología) se pudo tener una valoración de las posibilidades de fiabilidad de la metodología, no obstante se pueda implementar un flujo completo de la metodología, para tener un grado de seguridad mayor de las numerosas oportunidades y fortalezas con que puede contar la forma propuesta de analizar videos para detectar y caracterizar objetos de forma automática.

## **Publicaciones derivadas**

Durante la investigación llevada a cabo en relación con la tesis presentada, se realizaron las siguientes publicaciones y presentaciones:

### **Plataforma Libre de Monitoreo y Análisis de Medias**

Yanio Hernandez Heredia, Aneli Valdés Acosta, María de Dolores Guardia

*I Taller de Gestión y Procesamiento de Video y Sonido Digital. UCiencia 2010*

### **Modelando una metodología para la detección de objetos en sistemas de catalogación automática**

Yanio Hernandez Heredia, Aylin Estrada Velazco

*I Taller de Gestión y Procesamiento de Video y Sonido Digital. UCiencia 2010*

### **Sistema de Captura e Indexación de Videos**

Yanio Hernandez Heredia, Aylin Estrada Velazco, Jean Michel Suarez

*I Taller de Gestión y Procesamiento de Video y Sonido Digital. UCiencia 2010*

### **La Información en la televisión. Software para su control**

Yanio Hernandez Heredia

*IV Congreso Internacional de Tecnologías Contenidos Multimedia y Realidad Virtual. Informática 2009*

### **Sistema de Captura y Catalogación de Medias**

Yanio Hernandez Heredia, Aneli Valdés Acosta, María de Dolores Guardia

*XVII Fórum de Ciencia y Técnica. UCI 2009*

### **La Información en la televisión. Software para su control**

Yanio Hernandez Heredia, Aliosmi López Velázquez

*V Taller de Procesamiento Digital de imágenes y video. UCiencia 2008*

## Referencias bibliográficas

*Edición de video digital*. [En línea] [Citado el: 16 de Marzo de 2010.] <http://redcamelot.com/ediciondigital/diccionario.htm>.

**A., Lic. Verónica Tenorio. 2003.** Fundación CEPEC. [En línea] 2003. [http://www.fundacionsepec.cl/estudio/archivos/agosto/presentacion\\_Estructura\\_de\\_una\\_Clase.pdf](http://www.fundacionsepec.cl/estudio/archivos/agosto/presentacion_Estructura_de_una_Clase.pdf).

Adobe Developer Connection. *Video y la Web*. [En línea] [Citado el: 3 de Marzo de 2010.] [http://www.adobe.com/es/devnet/flash/articles/video\\_guide.html](http://www.adobe.com/es/devnet/flash/articles/video_guide.html).

**Akutsu, A. 1992.** *Video Indexing Using Motion Vectors*. Boston, MA, EE.UU : SPIE Conference on Visual Communications and Image, 1992.

**Alejandro Abejon, Ismael Mateos. 2009.** V Jornadas en Tecnología del Habla. [En línea] 2009. [http://jth2008.ehu.es/cd/pdfs/articulo/art\\_10.pdf](http://jth2008.ehu.es/cd/pdfs/articulo/art_10.pdf).

**Alfredo, M. 2008.** Vision AIBO. [En línea] 2008. [Citado el: 16 de 01 de 2010.] <http://cannes.itam.mx/Alfredo/Espaniol/Cursos/Robotica/Material/VisionAIBO.pdf>.

**Angulo, J. 2003.** *Morphologie mathématique at indexation d' images couleur*. s.l. : Tesis doctoral, 2003.

**Aquije, Jorge. 2009.** Iphala .NET. [En línea] 2009. [Citado el: 13 de 01 de 2010.] [http://www.wiphala.net/courses/intelligent\\_systems/ICSI271](http://www.wiphala.net/courses/intelligent_systems/ICSI271).

**B. Leibe, A. Leonardis, and B. Schiele. 2008.** *Robust object detection with interleaved categorization and segmentation*. s.l. : International Journal of Computer Vision, 2008. Vol 77 No 1.

**Bescós, J. 2004.** *Real-Time Shot Change Detection Over Online MPEG-2 Video*. [En línea] 04 de 2004. Vol 14 No 4 pp 475-484.

**Blanco, Jairo.** jairoblanco.com. *Tipos de Plataforma Web*. [En línea] [Citado el: 8 de Diciembre de 2009.] [http://www.jairoblanco.com/net/plataformas-web\\_90..](http://www.jairoblanco.com/net/plataformas-web_90..)

**2009.** Blog Terrier. *Lenguajes de Programación Orientado a Objeto*. [En línea] Mayo de 2009. [Citado el: 3 de Marzo de 2010.] <http://www.blogterrier.com.ar/2009/05/lenguajes-de-programacionorientada.html>.

**Breiman., L. 2001.** *Random forests*. s.l. : Machine Learning, 2001. 45(1):5–32.

**Brunelli, R. 1997.** *A Survey on the Automatic Indexing of Video Data (O. y Modena, C.M)*. EUA : Journal of Visual Communication and Image Representation, 1997.

**Cabezas, Raúl Trujillo. 2004.** *Aplicaciones del método Delphi*. Colombia , Bogota : s.n., 2004.

- Cabezas, Raúl Trujillo. 2004.** *Aplicaciones del método Delphi*. Colombia : Bogota, 2004.
- Campos, Jorge. 1999.** *Universidad de Mexico Grupo de Vision Artificial*. [En línea] 1999. [www-optica.inaoep.mx](http://www-optica.inaoep.mx).
- Chen, S. 2002.** *Scene Change Detection by Audio and Video Clues*. s.l. : IEEE Transactions on Speech and Audio Processing, 2002. Vol. II, pp. 365- 368.
- Cheriet, M. Suen, C. Y. 1998.** *A recursive thresholding technique for image segmentation*. s.l. : IEEE Transactions of image recognition, 1998.
- Cootes, T. 2003.** Active appearance models. s.l. : IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, Vol. 23.
- Coughlan, J. 2000.** Efficient deformable template detection and localization without user initialization. [En línea] 2000. [Citado el: 23 de 01 de 2010.] [www.eprints.kfupm.edu.sa/36761/1/36761.pdf](http://www.eprints.kfupm.edu.sa/36761/1/36761.pdf).
- D. Hoiem, S. Divvala, J. Hays. 2008.** University of Illinois at Urbana-Champaign, Beckman Institute. *A Unified Approach for Detection, Classification and Segmentation*. [En línea] 2008. [Citado el: 12 de 06 de 2009.] [http://www.ri.cmu.edu/publication\\_view.html?pub\\_id=6227](http://www.ri.cmu.edu/publication_view.html?pub_id=6227).
- de, Conocimiento Base.** Emotic. *Características de los flujos de videos*. [En línea] [Citado el: 21 de Noviembre de 2009.] <http://emotic.tv/ayuda/knowledge-base8.php>.
- Diccionario, Informático.** Alegsa.com. *Definición de FLV*. [En línea] [Citado el: 2009 de Diciembre de 2009.] <http://www.alegsa.com.ar/Dic/flv.php>.
- Elschlager, M. Fischler and R. 2005.** *Pictorial structures for object recognition*. s.l. : International Journal of Computer Vision, 2005. vol 61 No 1.
- Escolano, F. 2007.** Aprendizaje: Boosting and Adaboost. [En línea] 2007. [Citado el: 13 de 06 de 2006.] <http://www.cs.ucsd.edu/~yfreund/adaboost/index.html>.
- 2009.** esDebian. *Mencoder: Un poderoso codificador de video* . [En línea] Octubre de 2009. [Citado el: 6 de Marzo de 2010.] <http://franky32.esdebian.org/36418/mencoder-poderoso-codificador-video>.
- Estellés, Luis Alfaro Ferreres y María José Roca.** Formatos Gráficos-Sonidosvideos. [En línea] [Citado el: 5 de Diciembre de 2009.] <http://wwwdi.ujaen.es/asignaturas/informatica/Teoria/Software%20Grafico/Graficodoc/Graficos%20audio%20y%20video.doc>.
- Estructure. *estructuretv.com*. [En línea] [Citado el: 2009 de Diciembre de 2009.] <http://www.estructuretv.com/>.
- Everingham, M. 2007.** The PASCAL Visual Object Classes. [En línea] 2007. [Citado el: 18 de 12 de 2009.] <http://www.pascal-network.org/challenges/VOC/voc2007/workshop>.

**F. Schroff, A. Criminisi, and A. Zisserman. 2008.** *Object class segmentation*. British : BMVC, 2008.

—. **2008.** *Object class segmentation*. s.l. : BMVC, 2008.

**Felzenszwalb, Pedro. 2008.** Journal Publications/Books Ramanan. [En línea] 2008. [Citado el: 15 de 12 de 2009.] <http://www.ics.uci.edu/~dramanan/>.

**Fronc, Marisu.** CreativeCow.net. *Avanzadas de formatos de auditorías*. [En línea] [Citado el: 5 de Diciembre de 2009.] <http://geo.creativecow.net/es/a/170>.

**Furlán, Dr. Alfredo. 1999.** *Metodología de la Enseñanza*. México : s.n., 1999. Aportaciones a la didáctica de la Educación Superior.

**G. Ciocca, R. Schettini. 2006.** Dipartimento di Informatica Sistemistica e Comunicazione (DISCo). [En línea] 2006. [www.springerlink.com/index/D197140484867001.pdf](http://www.springerlink.com/index/D197140484867001.pdf).

**González, Y. Sarduy Domínguez y P. Urra.** bvs.sld.cu. *En busca de una plataforma ideal*. [En línea] [Citado el: 5 de Diciembre de 2009.] [http://bvs.sld.cu/revistas/aci/vol14\\_4\\_06/aci11406.htm](http://bvs.sld.cu/revistas/aci/vol14_4_06/aci11406.htm).

**Gopalakrishnam, P. S. and S. Chen. 1998.** *Speaker, environment and channel change detection and clustering via the Bayesian information criterion*. Lansdowne : DARPA Broadcast News Transcription and Understanding Workshop, 1998. pp. 127–132..

Guía Ubuntu. *Totem*. [En línea] [Citado el: 3 de Marzo de 2010.] <http://www.guiaubuntu.org/index.php?title=Totem>.

Hiparco. *Uso de Mencoder* . [En línea] [Citado el: 3 de Marzo de 2010.] <http://guimi.net/blogs/hiparco/uso-de-mencoder/>.

**Huttenlocher, D. 2004.** *Distance transforms of sampled functions*. s.l. : Cornell University CIS, 2004.

**J. Shotton, M. Johnson, and R. Cipolla. 2008.** *Semantic texton forests for image categorization and segmentation*. Francia : CVPR, 2008.

**Jack, Keith. 2002.** *Dictionary of Video and Television Technology*. Estados Unidos : Elsevier Science, 2002. 1-878707-99-X.

**K. Levi, Y. Weiss. 2004.** *Learning object detection from a small number of examples*. s.l. : Proceeding. Computer Vision and Pattern recognition, 2004.

**Keller, Frank. 2006.** *Introduction to Machine Learning. Connectionist and Statistical Language Processing*. Universitat des Saarlandes : Computerlinguistik, 2006.

Kioskea.net. *Reproducir un archivo en formato WMV* . [En línea] [Citado el: 5 de Marzo de 2010.] <http://es.kioskea.net/faq/526-reproducir-un-archivo-en-formato-wmv>.

**L. Van Gool, M. Everingham, C. Williams, A. Zisserman. 2006.** The PASCAL Visual Object Classes Challenge 2006 (VOC2006). [En línea] 2006. [Citado el: 15 de 01 de 2010.] <http://www.pascal-network.org/challenges/VOC/voc2006>.

**Lanzillotta, Analía.** MasterMagazine. *Archivo*. [En línea] [Citado el: 5 de Diciembre de 2009.] <http://www.mastermagazine.info/termino/3900.php>.

**Laptev, Ivan. 2005.** *On Space-Time Interest Points*. [En línea] 2005. [Citado el: 13 de 07 de 2009.] [http://www.irisa.fr/vista/Papers/2005\\_ijcv\\_laptev.pdf](http://www.irisa.fr/vista/Papers/2005_ijcv_laptev.pdf).

—. **2009.** histograms, Improving object detection with boosted. [En línea] 2009. [www.elsevier.com/locate/imavis](http://www.elsevier.com/locate/imavis).

**Lempitsky, Juergen Gall and Victor. 2009.** Class-Specific Hough Forests for Object Detection. [En línea] 2009. [Citado el: 15 de 03 de 2010.] [http://www.vision.ee.ethz.ch/~gallju/download/jgall\\_houghforest\\_cvpr09.pdf](http://www.vision.ee.ethz.ch/~gallju/download/jgall_houghforest_cvpr09.pdf).

**2007.** Linux Online . *What is Linux?* . [En línea] 2007. [Citado el: 6 de Febrero de 2010.] <http://www.linux.org/info/>.

**M. Fritz, B. Leibe, B. Caputo, B. Schiele. 2005..** *Integrating representative and discriminative models for object category detection*. Beijing, China : 10th International Conference on Computer Vision, , 2005.

**Macromedia, System. 2000.** Tutoriales de Software. [En línea] 2000. [Citado el: 12 de 06 de 2009.] <http://www.todoprogramas.com/manuales/ficheros/2008/1.6184.8104.doc>.

**Medrano, Raúl Igual y Carlos. 2008.** Tutorial de OpenCV. [En línea] 2008. [Citado el: 19 de 02 de 2010.] [http://docencia-eupt.unizar.es/ctmedra/tutorial\\_opencv.pdf](http://docencia-eupt.unizar.es/ctmedra/tutorial_opencv.pdf).

**2006.** Meth Blog. *Creación de video con Software Libre*. [En línea] 11 de Noviembre de 2006. [Citado el: 20 de Diciembre de 2009.] <http://meth.nireblog.com/post/2006/11/27/creacion-de-video-consoftware-libre..>

**MIT, The. 2004.** Introduction to Machine Learning. [En línea] 2004. [Citado el: 26 de 10 de 2009.] <http://www.cmpe.boun.edu.tr/~ethem/i2ml/>.

**Molina, Maria Pinto.** COMS. *Formatos*. [En línea] [Citado el: 5 de Diciembre de 2009.] <http://www.mariapinto.es/e-coms/formatos.htm>.

Noticias Tecnología. *Contenedores Multimedia* . [En línea] [Citado el: 16 de Marzo de 2010.] <http://www.uol.com.ar/noticias/tecnologia/datadelanet/20050328/nota3.html>.

**O'Connor, Edward. 2009.** [edward.oconnor.cx](http://edward.oconnor.cx). *Web platform: what it is*. [En línea] 21 de Mayo de 2009. [Citado el: 2009 de Diciembre de 2009.] <http://edward.oconnor.cx/2009/05/what-the-web-platform-is>.

**Otsu, N. 1979.** A threshold selection method from grey-level histograms. [En línea] 1979. [Citado el: 12 de 01 de 2010.] [http://ricic.cbnu.ac.kr/docu/2002year/3/2002\\_13.pdf](http://ricic.cbnu.ac.kr/docu/2002year/3/2002_13.pdf).

- P. Gil, F. Torres, F.G.Ortiz. 2004.** *Grupo de Automática, Robótica y Visión.* [En línea] 2004. [Citado el: 10 de 06 de 2009.] <http://www.ceafiac.es/actividades/jornadas/XXV/documentos/19-abtsufulva.pdf>.
- P. Viola, M. Jones. 2001.** *Rapid object detection using a boosted cascade of simple features.* Hawai : Proceeding Computer Vision and Pattern Recognition, 2001.
- Pedro Felzenszwalb, Ross Girshick, David McAllester and Deva Ramanan. 2009.** *Object Detection with Discriminatively Trained Part Based Models.* s.l. : IEEE, 2009.
- Pentland A., Picard R., Davenport G. and Haase K. 1995.** *Video and Image Semantics: Advanced Tools for Telecommunications.* s.l. : IEEE MultiMedia , 1995.
- Plasencia, Alberto Onildo.** Galeon.hispavista.com. *La Tecnología.* [En línea] [Citado el: 5 de Diciembre de 2009.] <http://galeon.hispavista.com/tecno/cienotecno.htm>.
- Porikli, F. 2005.** Integral histogram: a fast way to extract histograms in cartesian spaces. [En línea] 2005. [www.merl.com/papers/docs/TR2005-057.pdf](http://www.merl.com/papers/docs/TR2005-057.pdf).
- R. Cipolla, J. Shotton and M. Johnson. 2008.** *Semantic texton forests for image categorization and segmentation.* Alaska : CVPR, 2008.
- R. Fergus, P. Perona, and A. Zisserman. 2004.** *Object class recognition by unsupervised scale-invariant learning.* s.l. : IEEE Conference on Computer Vision and Pattern Recognition, 2004.
- Ramos, Juan Luis Bravo. 2005.** Secretariado de Recursos Audiovisuales y Nuevas Tecnologías. Universidad de Sevilla. *LOS MEDIOS DE ENSEÑANZA: CLASIFICACIÓN, SELECCIÓN Y APLICACIÓN.* [En línea] 2005. <http://www.sav.us.es/pixelbit/pixelbit/articulos/n24/n24art/art2409.htm>.
- Ramos, Rubén.** Linuca.org . *Formatos de video.* [En línea] [Citado el: 5 de Diciembre de 2009.] <http://www.linuca.org/body.phtml?nIdNoticia=148>.
- Resource, Animation. 1999.** The Animation Resource. [En línea] 1999. [Citado el: 15 de 06 de 2009.] <http://www.keyframeonline.com/>.
- Rodríguez, I. M. Aportela.** bvs.sld.cu. *Las tecnologías de información y comunicación en función de la organización.* [En línea] [Citado el: 5 de Diciembre de 2009.] [http://bvs.sld.cu/revistas/aci/vol16\\_4\\_07/aci041007.html](http://bvs.sld.cu/revistas/aci/vol16_4_07/aci041007.html).
- Rui Y., Huang T. S. and Mehrotra. 1999.** *Exploring Video Structure Beyond the Shots.* Texas, USA : Exploring Video Structure Beyond the Shots, 1999.
- Sánchez, Lic. Néstor Fernández. 2007.** Universidad Nacional Autónoma de México. [En línea] 2007. <http://www.tuobra.unam.mx/obrasPDF/publicadas/021123232113.html>.
- Sharp, T. 2008.** *Implementing decision trees and forests on a GPU.* s.l. : ECCV, 2008. pp. 595–608.

**Shotton, J. M. Winn and J. 2007.** *The layout consistent random field for recognizing and segmenting partially occluded objects.* s.l. : CVPR, 2007.

**Socarras., Lic. José Manuel Ruiz. 1999.** Universidad Autónoma de Madrid. *Los métodos de enseñanza en la Educación Superior Cubana.* [En línea] 1999. [http://www.uam.es/otros/rinace/biblioteca/documentos/Ruiz\(1999\).pdf](http://www.uam.es/otros/rinace/biblioteca/documentos/Ruiz(1999).pdf).

Softonic. *Completo servidor de multidifusión de audio y vídeo y conversor multiformato.* [En línea] [Citado el: 5 de Febrero de 2010.] <http://ffmpeg.softonic.com/linux>.

Softonic. *Completísimo reproductor de todo tipo de vídeos .* [En línea] [Citado el: 3 de Marzo de 2010.] <http://mplayer.softonic.com/linux>.

sonybiz.net. *SONAPS.* [En línea] [Citado el: 21 de Noviembre de 2009.] <http://www.sonybiz.net/sonaps>.

**Starópoli, Nicolás. 2006.** *Tratamiento digital de imágenes y visión artificial.* Italia, Genova : UNIVERSIDAD TECNOLÓGICA NACIONAL GRUPO DE INTELIGENCIA ARTIFICIAL (GIA), 2006.

**Stuart Andrews, Ioannis Tsochantaridis. 2003.** Support vector machines for multiple-instance learning. [En línea] 2003. [Citado el: 22 de 02 de 2010.] [www.citeseerx.ist.psu.edu](http://www.citeseerx.ist.psu.edu).

**Tenzer, Mario Simón. 2007.** *Archivos formatos y extensiones.* 2007.

**Tonomura Y., Akutsu A., Otsugi K., and Sadakata T. 1995.** *VideoMAP and VideoSpacecon: Tools for automatizing video content.* s.l. : ACM INTERCHI '95 Conference, 1995.

**Toruncha, Dr. José Zilberstein. 2000.** *Aprendizaje, Enseñanza y Desarrollo.* 2000.

**Triggs, N. Dalal and B. 2005.** *Histograms of oriented gradients for human detection.* EUA : IEEE Conference on Computer Vision, 2005.

**Tucker, Allen B. 1992.** *Lenguajes de Programación.* Mexico : McGraw-Hill , 1992.

**Turrof, H. Listone y M. 1975.** *The Delphi Method. Techniques and Applications.* Estados Unidos : s.n., 1975.

**Valverde, Jorge. 2006.** Escuela Académico Profesional de Informática. . [En línea] 2006. [Citado el: 21 de 12 de 2010.] [www.seccperu.org/files/DeteccióndeBordes-Canny.pdf](http://www.seccperu.org/files/DeteccióndeBordes-Canny.pdf).

Video LAN. *The cross-platform open-source multimedia framework, placer and server.* [En línea] [Citado el: 3 de Marzo de 2010.] <http://www.videolan.org/vlc/>.

**Y. Freund, R.E. Schapire., 1997.** *A decision-theoretic generalization of on-line learning and an application to boosting.* EUA : Journal of Computer and System Sciences, 1997. p. 119–139..

**Yingzi, D. 2004.** Unsupervised approach to color video thresholding. *Optical Engineering*, Vol. 43, No 2. [En línea] 2004. [Citado el: 12 de 01 de 2010.] <http://www.engr.iupui.edu/~yidu/pub.html>.

**Z, Chetverikov D. and Szabo. 1999.** *A Simple and Efficient Algorithm for Detection of High Curvature Points in Planar Curves*. Austria : 23rd Workshop of the Austrian Pattern Recognition, 1999.

**Zaiane, Osmar R. 2003.** Principles of Knowledge Discovery in Databases. [En línea] 2003. [Citado el: 28 de 10 de 2009.] <http://www.cs.ualberta.ca/~zaiane/courses/cmput690/slides/Chapter8/index.html>.

**Zhou, Y. Zhu and D. 2003.** *Scene Change Detection Based on Audio and Video Content Analysis*. s.l. : ICCIMA, 2003.

## Anexos

### Anexo 1: Algoritmo para la summarización de Videos

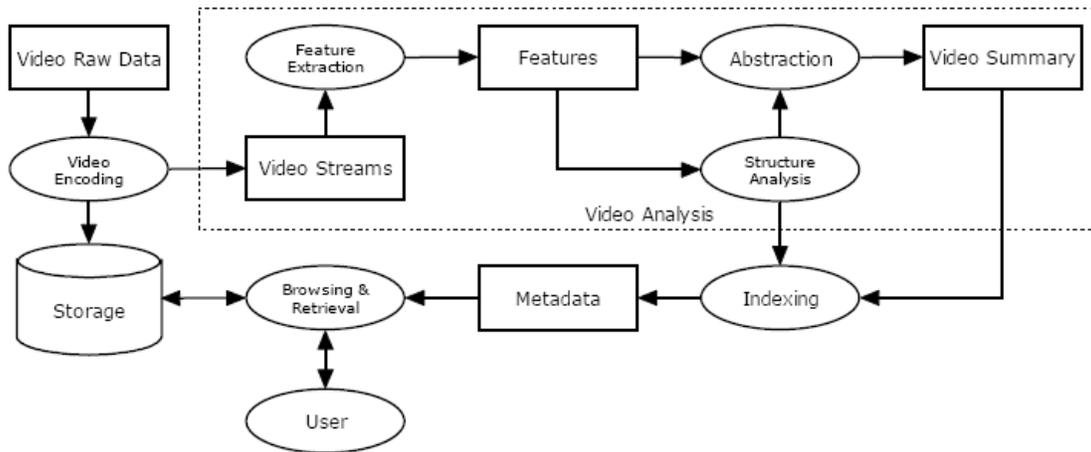


Figura 16. Aplicación General para tareas de análisis e indexación de videos. (G. Ciocca, 2006)

**Anexo 2: Comparación de métodos basados en bosques binarios y la transformada de Hough.**

<b>Methods</b>	<b>UIUC-Single</b>	<b>UIUC-Multi</b>
<i>Hough-based methods</i>		
Implicit Shape Model [10]	91%	–
ISM+verification [10]	97.5%	95%
Boundary Shape Model [17]	85%	–
<i>Random forest based method</i>		
LayoutCRF [27]	93%	–
<i>State-of-the-art</i>		
Mutch and Lowe CVPR'06 [15]	99.9%	90.6%
Lampert et al. CVPR'08 [9]	98.5%	98.6%
<i>Our approach</i>		
<b>Hough Forest</b>	98.5%	98.6%
HF - Weaker supervision	94.4%	–

Figura 17. Comportamiento de diferentes métodos en las dos Bases de Datos de coche de UIUC

### Anexo 3: Diagrama de clases de la técnica Bosques de Hough para la detección de objetos.

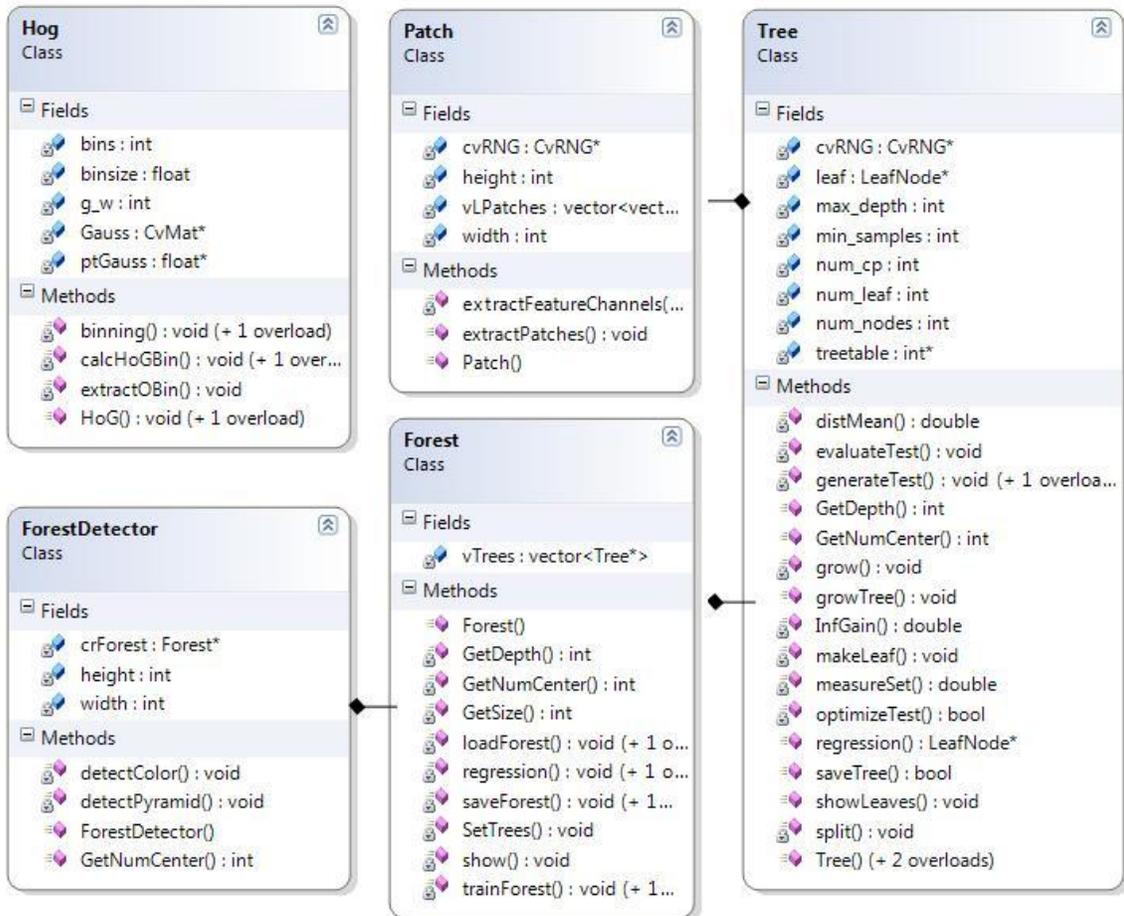


Figura 18. Diagrama de clases Forest Detector. Implementación en C++, compilada para Linux con OpenCV

#### Anexo 4: Aplicación de la técnica a dos imágenes de dos bases de datos de imágenes.

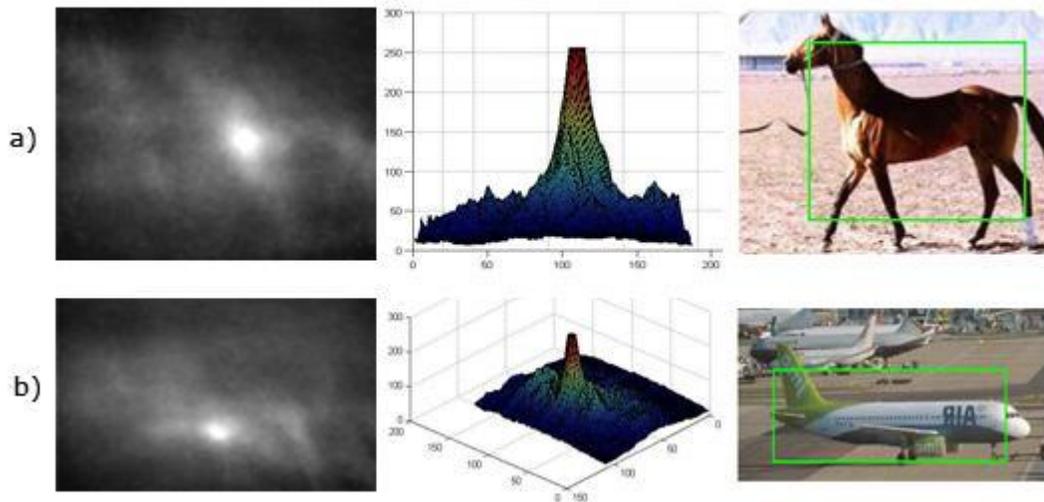


Figura 19. a) Weizzman Horse DB b) Airplanes Dates Trevick DB (Imagen de Hough, Gráfica de intensidad en el punto medio, imagen original con el Bounding box).

#### Anexo 5: Preguntas de la Encuesta a los Expertos.

Compañero(a):

Usted ha sido elegido, por su calificación y experiencia en su profesión, como experto para evaluar los resultados teóricos de esta investigación, por lo que el autor le pide que ofrezca sus ideas y criterios sobre beneficios o insuficiencias que pudiera presentar al ser aplicada la propuesta en los proyectos productivos del Departamento Señales Digitales. Sus criterios serán manejados de forma anónima y además le estamos agradeciendo su valiosa colaboración que contribuirá a mejorar esta metodología que será aplicada con el fin de incorporar valor agregado a los proyectos de la facultad.

Muchas gracias.

1. ¿Cuál es su relación con el tema de investigación expuesto en la tesis dentro de sus obligaciones laborales?

---

2. ¿Cómo usted valora la necesidad del empleo de la metodología propuesta dentro de los proyectos productivos del centro GEySED asociados al tema?

Alta\_\_\_                      Media\_\_\_                      Baja\_\_\_

3. ¿Cómo usted valora las posibilidades de aplicación de las técnicas propuestas en la metodología?

Alta\_\_\_                      Media\_\_\_                      Baja\_\_\_

4. ¿Cómo usted valora la implementación de la detección de objetos dentro de un sistema de gestión audiovisual?

Difícil\_\_\_      Media \_\_\_      Fácil \_\_\_\_\_

5. ¿Cómo usted valora la calidad de la investigación que permite detectar objetos en sistemas de catalogación semi-automática y automática de videos?

Alta\_\_\_                      Media\_\_\_                      Baja\_\_\_

6. Escriba las sugerencias que desee hacer para cualquiera de los procesos definidos en la metodología.

---

---

---

---

---

---

---